

Deterministic Chaos in General Relativity

Edited by

David Hobill

Adrian Burd and

Alan Coley

NATO ASI Series

Deterministic Chaos in General Relativity

NATO ASI Series

Advanced Science Institutes Series

A series presenting the results of activities sponsored by the NATO Science Committee, which aims at the dissemination of advanced scientific and technological knowledge, with a view to strengthening links between scientific communities.

The series is published by an international board of publishers in conjunction with the NATO Scientific Affairs Division

A	Life Sciences	Plenum Publishing Corporation
B	Physics	New York and London
C	Mathematical and Physical Sciences	Kluwer Academic Publishers
D	Behavioral and Social Sciences	Dordrecht, Boston, and London
E	Applied Sciences	
F	Computer and Systems Sciences	Springer-Verlag
G	Ecological Sciences	Berlin, Heidelberg, New York, London,
H	Cell Biology	Paris, Tokyo, Hong Kong, and Barcelona
I	Global Environmental Change	

Recent Volumes in this Series

Volume 325 — Statistical Mechanics, Protein Structure, and Protein Substrate Interactions
edited by Sebastian Doniach

Volume 326 — Linking the Gaseous and Condensed Phases of Matter:
The Behavior of Slow Electrons
edited by Loucas G. Christophorou, Eugen Illenberger,
and Werner F. Schmidt

Volume 327 — Laser Interactions with Atoms, Solids, and Plasmas
edited by Richard M. More

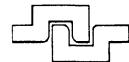
Volume 328 — Quantum Field Theory and String Theory
edited by Laurent Baulieu, Vladimir Dotsenko, Vladimir Kazakov,
and Paul Windey

Volume 329 — Nonlinear Coherent Structures in Physics and Biology
edited by K. H. Spatschek and F. G. Mertens

Volume 330 — Coherent Optical Interactions in Semiconductors
edited by R. T. Phillips

Volume 331 — Hamiltonian Mechanics: Integrability and Chaotic Behavior
edited by John Seimenis

Volume 332 — Deterministic Chaos in General Relativity
edited by David Hobill, Adrian Burd, and Alan Coley



Series B: Physics

Deterministic Chaos in General Relativity

Edited by

David Hobill

University of Calgary
Calgary, Alberta, Canada

and

Adrian Burd and Alan Coley

Dalhousie University
Halifax, Nova Scotia, Canada

Proceedings of a NATO Advanced Research Workshop on
Deterministic Chaos in General Relativity,
held July 25–30, 1993,
in Kananaskis, Alberta, Canada

Library of Congress Cataloging-in-Publication Data

Deterministic chaos in general relativity / edited by David Hobill and Adrian Burd and Alan Coley.

p. cm. -- (NATO ASI series. Series B, Physics : v. 332)

"Published in cooperation with NATO Scientific Affairs Division."
"Proceedings of a NATO Advanced Research Workshop on Deterministic Chaos in General Relativity, held July 25–30, 1993, in Kananaskis, Alberta, Canada"--T.p. verso.

Includes bibliographical references and index.

ISBN 978-1-4757-9995-8

1. General relativity (Physics)--Congresses. 2. Chaotic behavior in systems--Congresses. 3. Dynamics--Congresses. I. Hobill, David W. II. Burd, Adrian. III. Coley, A. A. IV. NATO Advanced Research Workshop on Deterministic Chaos in General Relativity (1993 : Kananaskis, Alta.) V. Series.

QC173.6.D48 1994

530.1'1'01185--dc20

94-30691

CIP

Additional material to this book can be downloaded from <http://extras.springer.com>.

ISBN 978-1-4757-9995-8
DOI 10.1007/978-1-4757-9993-4

ISBN 978-1-4757-9993-4 (eBook)

© 1994 Springer Science+Business Media New York
Originally published by Plenum Press, New York in 1994
Softcover reprint of the hardcover 1st edition 1994.

All rights reserved

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without written permission from the Publisher

PREFACE

Nonlinear dynamical systems play an important role in a number of disciplines. The physical, biological, economic and even sociological worlds are comprised of complex nonlinear systems that cannot be broken down into the behavior of their constituents and then reassembled to form the whole. The lack of a superposition principle in such systems has challenged researchers to use a variety of analytic and numerical methods in attempts to understand the interesting nonlinear interactions that occur in the World around us.

General relativity is a nonlinear dynamical theory *par excellence*. Only recently has the nonlinear evolution of the gravitational field described by the theory been tackled through the use of methods used in other disciplines to study the importance of time dependent nonlinearities. The complexity of the equations of general relativity has been (and still remains) a major hurdle in the formulation of concrete mathematical concepts. In the past the imposition of a high degree of symmetry has allowed the construction of exact solutions to the Einstein equations. However, most of those solutions are nonphysical and of those that do have a physical significance, many are often highly idealized or time independent.

The vigorous interest in the chaotic behaviour of dynamical solutions to the Einstein equations has been helped by the development of numerical relativity. A few years ago several groups began, almost simultaneously, to examine numerically the behaviour of the Mixmaster model and found apparent discrepancies with the results that had been obtained a decade earlier. Subsequently other strong field problems in relativity were studied, again through the use of numerical methods, and these have begun to indicate that other unforeseen consequences of the nonlinear nature of general relativity may be present.

In order to gather the different studies and alternative approaches together, we felt that it would be useful to conduct a small workshop of researchers and students interested in this field of work. The resulting meeting was organized as part of the NATO Advanced Research Workshop Program and was held between July 25th and July 30th 1993 in Alberta's Kananaskis region located in the front ranges of the Canadian Rockies. The contributions in this volume provide a record of the presentations and

discussions held during that week. In addition a few *in absentia* contributions have been added since they also played important roles in stimulating discussions during the meeting.

We had two main aims in holding this meeting. The first was to provide a forum for participants to assess the current status of our understanding of chaotic behaviour in general relativity. In doing so we hoped to provide an atmosphere where the direction of future research effort could be discussed and long term collaborations could be organized. Our other aim was to provide a reference for both those currently active in the area and also for those who wished to enter the field. To this end we enlisted the involvement of researchers whose main area of study was dynamical systems theory; this also produced some very interesting new avenues of research, some of which are presented in the contributions contained within. We hope that those articles which are more pedagogical in nature will make the volume self contained and useful to students and active researchers alike.

These aims are also reflected in the goals of the NATO Advanced Research Workshop program; to review the current state of research in a particular field and provide a forum for debate on possible future directions of research. To this end we held several informal discussions as well as a more structured panel discussion¹. What follows is a précis of these discussions along with interesting comments raised during the workshop.

Much of the recent research on the Bianchi IX models has concentrated on demonstrating conclusively the existence of chaos in these systems. As was repeatedly pointed out during the meeting (see contributions by Francisco and Matsas, Burd, Berger and Rugh) the Lyapunov exponent, which is most often used in this context, is an inappropriate measure to use in General relativity. This has led to searches for gauge invariant measures of chaos which can be used when considering relativistic systems. It was argued by Berger and Misner that to say the Bianchi IX model is, or is not, chaotic is largely a matter of semantics. What is more important is to determine the consequences of this type of behaviour at physically relevant length scales. A suitable analogy might be the dynamical behaviour of the solar system. It is known that over time scales relevant to the evolution of the solar system, the orbits of certain planets are chaotic. However, on the time scale of human civilization, this behaviour is unimportant.

In the case of the oscillations occurring in the Bianchi IX cosmologies (a topic that received much discussion during the workshop) it appears that only a small finite number of these oscillations occur on length and time scales greater than the Planck length and time. It may be that arguments based upon the dynamics of a possible infinite number of oscillations occurring within physically unmeasurable scales may have as much relevance as those medieval philosophical arguments concerning the number of angels capable of dancing on the head of a pin.

Whilst the development of gauge invariant indicators of chaos is important in its own right, a great deal of work needs to be done to understand how the nonlinear behaviour of these models affects the physics. In this vein, the contributions by Choptuik, Berger *et al*, Bombelli, Calzetta and Tavakol are first steps in this direction.

Much of the analytic work for the Mixmaster model has involved varying degrees of approximation. In making these approximations some information about the full dynamical system is discarded and it remains unclear how important this is. Berger,

¹The panel consisted of Charles Misner, Beverly Berger, Matt Choptuik, Esteban Calzetta and Reza Tavakol.

Misner and Tavakol felt that the relationship between the various approximations used and the numerically generated flows remained unclear.

One of the problems with the Mixmaster model is that its phase space is noncompact. Churchill emphasized that a rigourous demonstration of chaotic behaviour may involve showing that the phase space of the dynamics is compact. It was suggested by Misner that it would be interesting to explore in more detail the possibility of finding a system of variables in which this was so.

Several contributions showed that chaos is by no means the only manifestation of the non-linear nature of general relativity and that these other effects, such as scaling phenomena in black hole formation, the development of spatial structure near the singularity and the use of fractals, should be investigated more.

Much of the research presented at this workshop relied heavily on numerical techniques. Choptuik and Hobill emphasized that the use of computer models will provide an increasingly important tool for investigating general relativity, particularly where the fields are strong. It was suggested that existing numerical codes could be made available in the form of a "toolbox" for research in the area.

Calzetta raised the analogy with current discussions in quantum cosmology. The full Einstein field equations form an infinite dimensional dynamical system and when studying nonlinear effects in general relativity one should be aware that one is really studying a finite dimensional version of the full system. This was echoed by Tavakol who drew the analogy with fluid dynamics where a considerable amount of research is done without looking for systems with exact symmetries. In pursuing this analogy, Tavakol noted that chaotic behaviour in fluids occurs typically with non-hamiltonian fluids whereas in general relativity, most cases studied so far deal with purely hamiltonian systems. The difficulty of examining models which have reduced symmetry is that one must consider numerical techniques. It was felt strongly that much work could be done by using toy models rather than dealing with the full system of Einstein equations, and reference was made to the work of Roger Ove.²

More general issues were also covered, such as questions of structural stability and genericity. Tavakol suggested that chaos within the Bianchi IX models should be examined to see if chaos was a generic feature or not. This may be difficult to implement. Even in the case of simpler systems, an answer is still lacking. More recently there have been arguments that the Bianchi IX cosmologies may be integrable (see the contribution of Contopoulos, *et al*). If these indications are correct the answer to the problem may not be far away.

Overall, the workshop accomplished many of its goals and more recently, interest in the results of the meeting has extended beyond the rarefied realm of the specialized researcher³. For those interested in dynamical general relativity and/or chaotic systems it is hoped that the contributions to this volume will be as stimulating and thought provoking to our readers as they were to those who had the opportunity to participate in the workshop.

As with all such endeavours, there is a long list of people who contributed time and effort to make the meeting run as smoothly as possible. We first acknowledge the Scientific Affairs Division of the North Atlantic Treaty Organization and the University of Calgary's Office of Research for the financial support they provided. Secondly we

²"Properties of a Model Einstein Equation", *Gen. Rel. and Grav.*, **22**, 631, (1990).

³See "Gravity's Chaotic Future" by Ivars Peterson in *Science News*, **144**, 369, (Dec. 4, 1993).

thank Grace LeBel and her staff at the Kananaskis Center for Environmental Research for providing a relaxed comfortable environment which enhanced the productivity of the workshop. The help of students, Teviet Creighton, Alistair Fraser, Robert van den Hoogen and Hossein Abolghasem made significant contributions to the organization of the meeting. We would also like to thank Patricia Vann at Plenum for her patience and good humour during the production of this volume.

The non-local members of the committee (Berger, Burd, Coley and Tavakol) would like to extend their thanks to David Hobill who, by virtue of being the only on-site member of the committee, bore the brunt of the workload in organising the meeting. Finally thanks go to the participants themselves who throughout the workshop and through their contributions to these proceedings demonstrate the excitement associated with attempting to understand the nature of deterministic chaos in general relativity.

Beverly Berger
Adrian Burd
Alan Coley
David Hobill
Reza Tavakol

CONTENTS

A Brief Review of “Deterministic Chaos in General Relativity”	1
D. W. Hobill	

MATHEMATICAL PRELIMINARIES

Introduction to Dynamical Systems	19
J. Wainwright	
A Short Course on Chaotic Hamiltonian Systems	63
R. C. Churchill	
Geometry of Perturbation Theory	89
R. Cushman	
Between Integrability and Chaos	103
D. L. Rod and R. C. Churchill	
On Defining Chaos in the Absence of Time	107
R. C. Churchill	
On the Dynamics of Generators of Cauchy Horizons	113
P. T. Chruściel and J. Isenberg	

COMPACT RELATIVISTIC SYSTEMS

Chaos in the Case of Two Fixed Black Holes	129
G. Contopoulos	

Particle Motion Around Perturbed Black Holes: The Onset of Chaos	145
L. Bombelli	

Critical Behaviour in Scalar Field Collapse	155
M. W. Choptuik	

COSMOLOGICAL SYSTEMS

Relativistic Cosmologies	179
M. A. H. MacCallum	

Homoclinic Chaos in Relativistic Cosmology	203
E. Calzetta	

Mixing Properties of $k = -1$ FLRW Models	237
G. F. R. Ellis and R. Tavakol	

Classical and Quantum Chaos in Robertson-Walker Cosmologies	251
R. Tomaschitz	

Relativistic Fractal Cosmologies	269
M. B. Ribeiro	

Self-Similar Asymptotic Solutions of Einstein's Equations	297
A. A. Coley and R. J. van den Hoogen	

Nonlinearly Interacting Gravitational Waves in the Gowdy T^3 Cosmology	307
B. K. Berger, D. Garfinkle, V. Moncrief and C. M. Swift	

BIANCHI IX (MIXMASTER) DYNAMICS

The Mixmaster Cosmological Metrics	317
C. W. Misner	

The Belinskii-Khalatnikov-Lifshitz Discrete Evolution as an Approximation to Mixmaster Dynamics	329
B. K. Berger	

How Can You Tell if the Bianchi IX Models Are Chaotic?	345
A. Burd	

The Chaoticity of the Bianchi IX Cosmological Model	355
G. Francisco and G. E. A. Matsas	

Chaos in the Einstein Equations-Characterization and Importance?	359
S. E. Rugh	

Integrability of the Mixmaster Universe	423
G. Contopoulos, B. Grammaticos, and A. Ramani	
Continuous Time Dynamics and Iterative Maps of Ellis-MacCallum-Wainwright Variables	433
T. D. Creighton and D. W. Hobill	
A Dynamical Systems Approach to the Oscillatory Singularity in Bianchi Cosmologies	449
P. K-H. Ma and J. Wainwright	
Participants	463
Index	467

A BRIEF REVIEW OF “DETERMINISTIC CHAOS IN GENERAL RELATIVITY”

David Hobill

Dept. of Physics and Astronomy, University of Calgary
Calgary, Alberta, T2N 1N4, Canada

Abstract. A short review of chaotic behavior and a dynamical systems approach to time-dependent solutions to the Einstein equations is provided in order to give an overview of the problems that have been studied and the approaches used to understand them.

1. INTRODUCTION

The time dependent behavior of the full nonlinear theory of general relativity presents a challenge to physicists and mathematicians alike. If general relativity has anything to say about our physical universe then we need to understand the implications of the theory and the predictions that it makes concerning the behavior of the world around us. On the mathematical side, the nonlinear structure of the theory (particularly its time dependence) has yet to be understood in a formal sense. Given that some of the simplest nonlinear dynamical systems can have very complicated (even stochastic) behavior in spite of the fact that the equations are deterministic, it is important to understand what such a knowledge means for general relativity.

For the past two to three decades the realization that the nonlinear time dependent behavior of certain nonlinear theories have an extreme sensitivity on initial conditions (i.e. chaos) has changed the way in which one views the evolution of physical, biological, and even sociological systems. No longer does one assume that our theories are sufficient to determine the time dependent behavior of the Universe if we were privileged to know what the initial conditions were.

This workshop represents the first time that a number of researchers in the field of relativity together with some mathematicians interested in dynamical systems have

met together to discuss the problems that need to be solved in order to understand the behavior of the Einstein equations in the strongly nonlinear regime.

The Einstein equations like the Navier-Stokes equations in hydrodynamics are nonlinear partial differential equations (PDE's) and since they have an infinite number of degrees of freedom, the simplest approach to understanding the nonlinear dynamical behavior of the equations is to reduce the number of degrees of freedom to a finite value. This is one reason why the cosmological principle is so popular from a mathematical (in addition to a physical) point of view; it reduces the gravitational degrees of freedom to one (the cosmological scale factor) in Friedmann-Lemaître-Robertson-Walker (FLRW) cosmologies.

Homogeneous but anisotropic cosmologies allow for the existence of more degrees of freedom and the nonlinear coupling of the variables that govern the behavior of anisotropic (but spatially homogeneous) cosmologies can result in interesting time dependent behavior. The homogeneous cosmologies also have the advantage that the Einstein equations reduce to a set of coupled nonlinear ordinary differential equations (ODE's) for which there exists an abundance of mathematical machinery for studying their properties. The studies of homogeneous cosmologies from a dynamical systems approach has become quite popular recently, especially with the increasing interest in chaotic nonlinear dynamical systems.

Another general relativistic problem that provides a set of ODE's capable of being studied through a dynamical systems approach is one that attempts to understand geodesic motion in a given spacetime. It is well known that Newtonian many-body systems are chaotic and therefore it is likely that a similar situation exists in general relativity. One attempt to understand this problem is to reduce the mass of one of the bodies and treat it as a test particle. Its motion will be then governed by the geodesic equations which may or may not be chaotic. Particularly interesting problems dealing with geodesic motion in black hole spacetimes have been and continue to be studied. Bombelli and Calzetta [1] have analyzed the behavior of geodesics in perturbed Schwarzschild spacetimes using a Melnikov method while Contopoulos [2] has examined the geodesics in a spacetime consisting of two Reissner-Nordstrom black holes in static equilibrium.

Purely gravitational systems with high degrees of symmetry do not in general have enough degrees of freedom to produce complicated nonlinear effects. However the number of degrees of freedom may be increased when non-gravitational fields are added to the right-hand-side of Einstein's equations. One example of such a system is the coupled gravitational and scalar field interaction that occurs in spherically symmetric spacetimes. Choptuik has recently found that there is some interesting nonlinear behavior in this system as the amplitude of the scalar field nears that required for collapse to a black hole [3].

Studies of the standard FLRW models and its nonlinear coupling to other fields provides another example of a case where more degrees of freedom are introduced and chaotic behavior becomes possible. Examples of these types of problems are those studied by Calzetta and his collaborators [4]. See also Calzetta's contribution to this volume). An interesting example is provided by a system in which a scalar field is conformally coupled to the gravitational field of a closed FRLW cosmology whose metric is given by:

$$ds^2 = a^2(\eta)[-d\eta^2 + d\chi^2 + \sin^2 \chi(d\theta^2 + \sin^2 \theta d\varphi^2)]$$

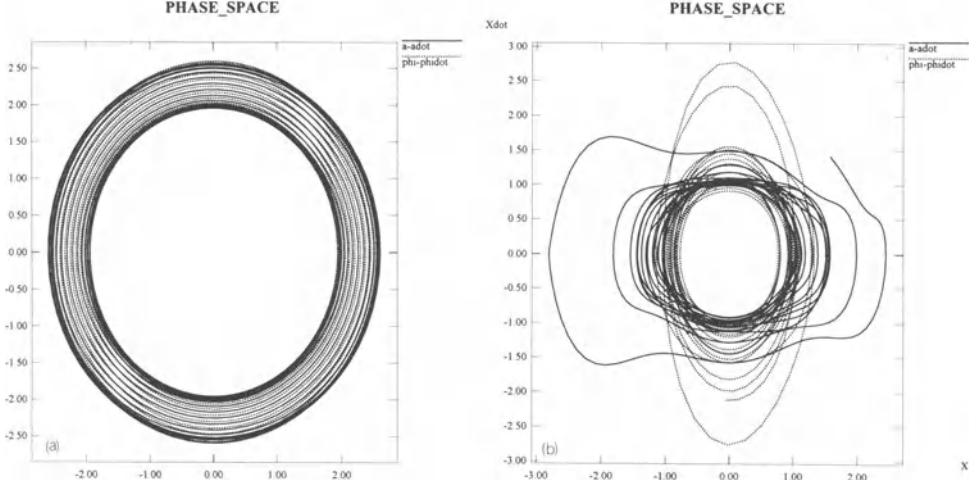


Figure 1. Phase space diagrams for the (a) $m = 0.1$ and (b) $m = 1.2$ scalar field and the expansion factor for a closed RW universe in conformal coordinates.

where η is the “conformal time” coordinate, χ is a periodic radial coordinate ($r = \sin \chi$), θ, φ are the standard 2-sphere angular coordinates and $a(\eta)$ is the scale factor.

The Einstein field equations and the Klein-Gordon equations can be derived from a Hamiltonian and written in terms of the scale factor a , its conjugate momentum p_a and a massive scalar field ϕ (having a mass m) and its conjugate momentum p_ϕ . The dynamical equations then become:

$$\begin{aligned}\dot{a} &= -p_a \\ \dot{p}_a &= a(1 - m^2\phi^2) \\ \dot{\phi} &= p_\phi \\ \dot{p}_\phi &= -\phi(1 + m^2a^2)\end{aligned}$$

where the overdot represents the derivative with respect to the conformal time coordinate. This system is closely related to a system of nonlinearly coupled harmonic oscillators and does provide an example of chaotic behavior when the mass of the scalar field reaches a critical value. Figure 1 provides an example of the phase space evolution of the scale factor and the scalar field in the chaotic regime and it is clear that the periodicity of the low amplitude solutions breaks down.

While general relativistic cosmologies provide “simple” systems, the problem associated with initial conditions here may be somewhat academic in that one is really asking questions about the behavior of a family of universes (many of which may not be physically significant).

On the other hand the sensitivity of initial conditions may also hold some secrets about the evolution of isolated systems in general relativity and the formation of complex structures in both space and time. This is a much harder question to deal with and is closely related to the existence of turbulence and pattern formation in hydrodynamics and plasma physics.

This review will introduce the methods and formalisms used to study Bianchi-IX (Mixmaster) spatially homogeneous spacetimes. No attempt is made to be complete nor will alternative gravity theories (e.g. higher dimensional theories, non-symmetric metrics, Cartan theory, etc.) which also have been studied from a dynamical systems approach.

2. BIANCHI IX or MIXMASTER COSMOLOGIES

The Bianchi cosmologies are a class of spatially homogeneous spacetimes having a particular isometry group structure on the spatial hypersurfaces [5, 6]. In the case of the Bianchi IX models the group is the $S0(3)$ (or $SU(2)$) rotation group. The metric for the spacetime in such a case can be given as:

$$ds^2 = -dt^2 + \eta_{ik}\omega^i\omega^k \quad (1)$$

where $\omega^i(i = 1, 2, 3)$ are one-forms determined by the structure constants of the group of motions and $\eta_{ik} = \text{diag}(a^2(t), b^2(t), c^2(t))$. The time parameter is considered to be a cosmological time that is constant on each spatial hypersurface. The time dependent functions a , b and c can be considered as the cosmological scale factors in the different directions of anisotropy.

For the Bianchi-IX models the frame vectors written in terms of the rotation angles $0 \leq \psi \leq 4\pi$, $0 \leq \theta \leq \pi$, $0 \leq \phi \leq 2\pi$ are:

$$\begin{aligned} \omega^1 &= \cos \psi d\theta + \sin \psi \sin \theta d\phi \\ \omega^2 &= \sin \psi d\theta - \cos \psi \sin \theta d\phi \\ \omega^3 &= d\psi + \cos \theta d\phi \end{aligned} \quad (2)$$

For these spacetimes the 3-volumes can be easily shown to be

$$V = 16\pi^2 abc.$$

In the special case of an isotropic universe $a = b = c = R/2$ and the metric reduces to the **closed FLRW** spacetime.

$$ds^2 = -dt^2 + R^2(t)d\Omega^2.$$

A number of approaches have been taken to study the time dependence of the Einstein equations derived from the metric (1) and these will be briefly discussed below. Some work still remains to be accomplished to show the equivalence of the different approaches. This is particularly true where approximations have been made based upon the behavior of certain terms appearing in the various formalisms.

3. THE BKL APPROACH

A straight forward approach was initiated by Belinskii, Khalatnikov and Lifshitz [7, 8, 9] (BKL) in an attempt to understand the generic behavior of general relativity in the vicinity close to a spacetime singularity. The idea to use anisotropic cosmologies was motivated by the assumption that in a small enough region, anisotropies would appear in

spatially dependent spacetimes and one could approximate such a small region with the Bianchi-IX structure and evolution. Some of these beliefs are still held today, although they remain unjustified.

The field equations for the metric (1) can be determined explicitly in the following form:

$$\begin{aligned} R_1^1 &= \frac{(\dot{abc})}{abc} + \frac{1}{2a^2b^2c^2}[a^4 - (b^2 - c^2)^2] = 0 \\ R_2^2 &= \frac{(\dot{bca})}{abc} + \frac{1}{2a^2b^2c^2}[b^4 - (c^2 - a^2)^2] = 0 \\ R_3^3 &= \frac{(\dot{cab})}{abc} + \frac{1}{2a^2b^2c^2}[c^4 - (a^2 - b^2)^2] = 0 \end{aligned}$$

and

$$R_0^0 = \frac{\ddot{a}}{a} + \frac{\ddot{b}}{b} + \frac{\ddot{c}}{c} = 0$$

Introducing an extrinsic curvature tensor given by $\text{diag}(f(t), g(t), h(t))$ the system of equations given above can be written in the following first order dynamical systems form:

$$\begin{aligned} \frac{da}{dt} &= -\frac{f}{a}, & \frac{df}{dt} &= f\left(\frac{g}{b^2} + \frac{h}{c^2} - \frac{f}{a^2}\right) + \frac{1}{2}\left(\frac{a^4}{b^2c^2} - \frac{b^2}{c^2} - \frac{c^2}{b^2} + 2\right) \\ \frac{db}{dt} &= -\frac{g}{b}, & \frac{dg}{dt} &= g\left(\frac{f}{a^2} + \frac{h}{c^2} - \frac{g}{b^2}\right) + \frac{1}{2}\left(\frac{b^4}{a^2c^2} - \frac{a^2}{c^2} - \frac{c^2}{a^2} + 2\right) \\ \frac{dc}{dt} &= -\frac{h}{c}, & \frac{dh}{dt} &= h\left(\frac{f}{a^2} + \frac{g}{b^2} - \frac{h}{c^2}\right) + \frac{1}{2}\left(\frac{c^4}{a^2b^2} - \frac{b^2}{a^2} - \frac{a^2}{b^2} + 2\right) \end{aligned}$$

Combining the equations above with the equation $R_0^0 = 0$ leads to the scalar constraint equation which acts as a first integral for the evolution equations given above:

$$2\left(\frac{gh}{b^2c^2} + \frac{fh}{a^2c^2} + \frac{fg}{a^2b^2}\right) - \frac{1}{2}\left(\frac{c^2}{a^2b^2} + \frac{b^2}{a^2c^2} + \frac{a^2}{b^2c^2}\right) + \frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2} = 0.$$

The time dependence (from the initial to the final singularity) of the scale factors a , b , and c are shown below in Figure 2.

The behavior that BKL wished to study was how the gravitational field evolved near the “big bang” and the “big crunch” singularities shown in Figure 2. For these particular times the “cosmic time” t was incapable of exploring these regions. BKL then introduced a new time coordinate τ by $d\tau = dt/abc$. Since the intervals of τ become infinitely large as the spatial volume (abc) goes to zero, this time parameterization is an example of one that “avoids” the singularity and is useful for exploring the behavior close to the singularity.

In addition to introducing a new time variable, BKL also introduced new metric coefficients $\alpha(t)$, $\beta(t)$ and $\gamma(t)$ defined by $a = e^\alpha$, $b = e^\beta$, and $c = e^\gamma$. The evolution equations simplify to:

$$\frac{d^2\alpha}{d\tau^2} = \frac{1}{2}[(e^{2\beta} - e^{2\gamma})^2 - e^{4\alpha}] \quad (3a)$$

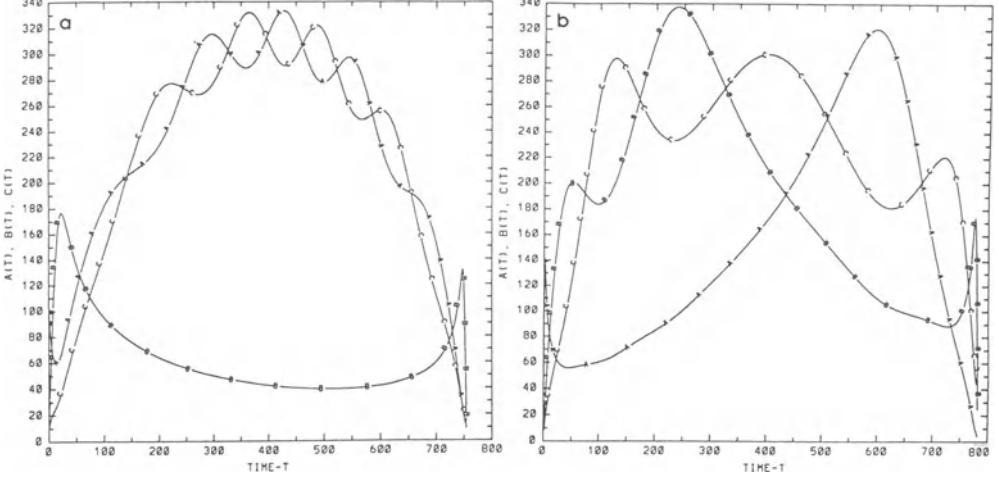


Figure 2. The evolution of the scale factors $a(t)$, $b(t)$ and $c(t)$ for an initially expanding vacuum Bianchi IX cosmology with unit lapse function. The initial values for (a) are: $a(0) = 100.$, $\dot{a}(0) = 10.$, $b(0) = 10.$, $\dot{b}(0) = 10.$, and $c(0) = 10.$ with $\dot{c}(0) = 1.2727$ determined from a solution of the Hamiltonian constraint. In (b) the initial values are: $a(0) = 10.$, $\dot{a}(0) = 100.$, $b(0) = 10.$, $\dot{b}(0) = 10.$, $c(0) = 10.$ and $\dot{c}(0) = -9.0977.$ Numerical methods are incapable of following the oscillations near the final singularity without a redefinition of the time parameter.

$$\frac{d^2\beta}{d\tau^2} = \frac{1}{2}[(e^{2\gamma} - e^{2\alpha})^2 - e^{4\beta}] \quad (3b)$$

$$\frac{d^2\gamma}{d\tau^2} = \frac{1}{2}[(e^{2\alpha} - e^{2\beta})^2 - e^{4\gamma}]. \quad (3c)$$

Introducing the derivatives of the metric components as the auxiliary variables $\kappa(\tau) = d\alpha/d\tau$, $\mu(\tau) = d\beta/d\tau$ and $\nu(t) = d\gamma/dt$ the the constraint equation becomes:

$$\frac{2[e^{2(\gamma+\beta)} + e^{2(\alpha+\gamma)} + e^{2(\alpha+\beta)}] - e^{4\alpha} - e^{4\beta} - e^{4\gamma}}{4} + \kappa\mu + \kappa\nu + \mu\nu = 0. \quad (3d)$$

Crucial to the understanding of the behavior of the Mixmaster dynamics is the Kasner solution. This is the closed form solution to the Bianchi I model for which the RHS's of Eqs(3) vanish. This means that α , β and γ are linear functions of τ (more specifically $\alpha = p_1\tau$, $\beta = p_2\tau$, $\gamma = p_3\tau$) with slopes determined by a single parameter, $1 \leq u \leq \infty$ such that $\Sigma p_i = \Sigma p_i^2 = 1$:

$$p_1 = \frac{-u}{1+u+u^2}, \quad p_2 = \frac{1+u}{1+u+u^2}, \quad p_3 = \frac{u(1+u)}{1+u+u^2}. \quad (4)$$

BKL were able to demonstrate that a sequence of Kasner solutions parametrized by u can provide a good approximation to the evolution close to a cosmological singularity. This can be justified by the smallness of the RHS's of Eqs. (3-c) when the singularity is approached.

Figures 3(a-d) show a typical evolution from initial data that satisfies the constraint equation and demonstrates the “piece-wise” linear behavior of the logarithmic scale factors α , β , and γ as functions of the time parameter τ .

There is an obvious Kasner type behavior for these scale factors and the Kasner regime lasts as long as one of the log(scale factors) remains less than zero. When the increasing scale factor reaches zero another Kasner regime takes over. Notice that the

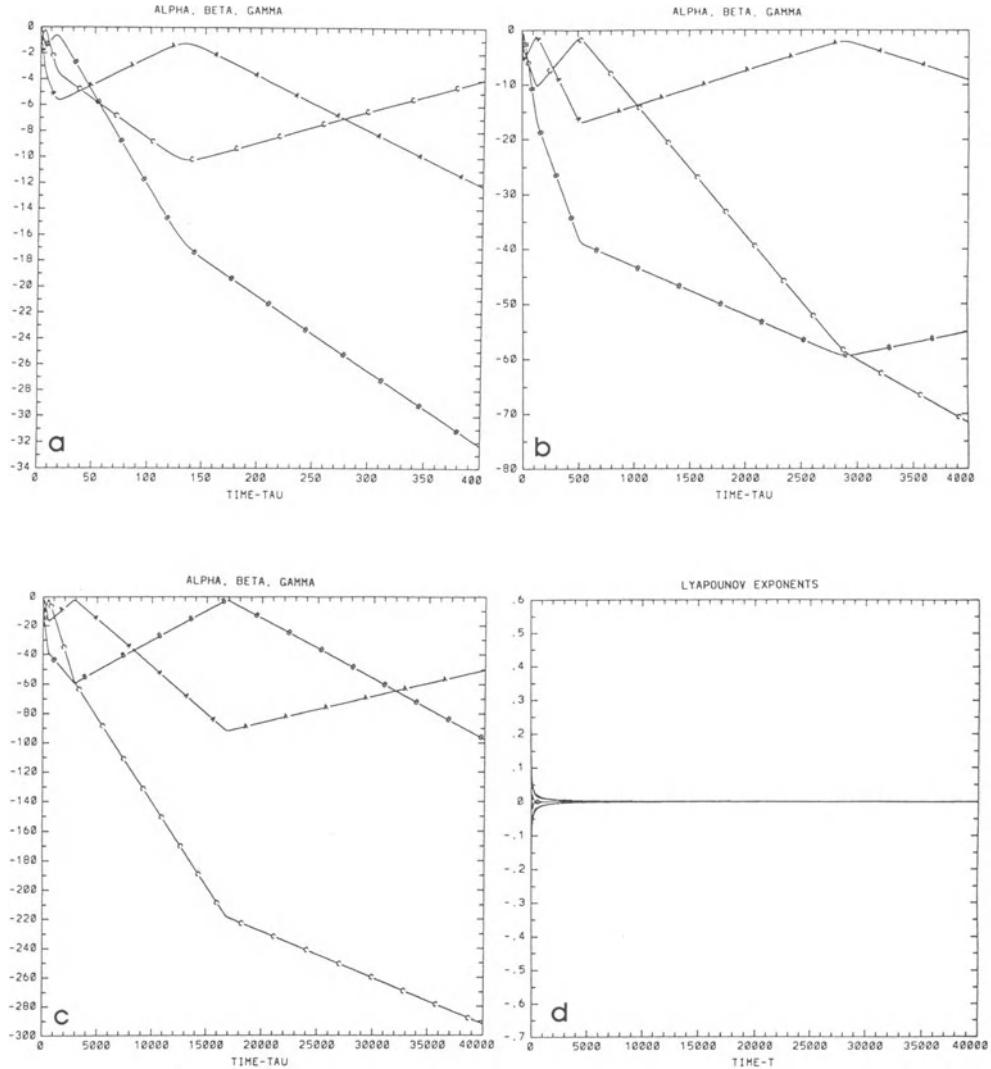


Figure 3. A typical numerical solution of the logarithm of the scale factors. (the curves are represented by A: α , B: β , and C: γ). The initial conditions for all four figures are: $\alpha(0) = 0.$, $\kappa(0) = -1.$, $\beta(0) = 0.$, $\mu(0) = .1$, $\gamma(0) = -1.$ and $\nu(0) = 3.4174 \times 10^{-2}$. The behavior of the scale factors over three orders of magnitude in the coordinate τ is shown in (a-c). Note that the period of the piecewise linear oscillations always increases in the direction of increasing τ . This is due to the choice of lapse function. A plot of the entire spectrum of Lyapunov exponents which approaches zero is shown in (d).

factor with the most negative value has a slope that becomes less and less negative as the Kasner regimes change. Eventually that scale factor has a positive slope and it will be the one that governs the next change in Kasner regimes when its value becomes equal to zero.

Each oscillation between two scale factors while one scale factor monotonically decreases is called an “epoch” (i.e. a Kasner regime) while the change of the originally monotonically decreasing factor from a negative to positive slope marks the beginning of an “era”. Note that an era may contain one or more Kasner epochs.

In terms of the parameter u that determines a particular Kasner solution, BKL found a simple transition rule for the change from one epoch to another. This can be written in the following discrete mapping:

$$u_{n+1} = u_n - 1 \quad \text{when } u_n \geq 2$$

$$u_{n+1} = \frac{1}{u_n - 1} \quad \text{when } 1 < u_n < 2$$

For example the following sequence describes the transition from one epoch to another starting from $u_0 = 3.5183\dots$

$$3.5183\dots \rightarrow 2.5183\dots \rightarrow 1.5183\dots$$

one then has a new era with only one epoch:

$$\frac{1}{.5183\dots} \rightarrow 1.0759\dots$$

which leads to another era

$$\frac{1}{.0759\dots} \rightarrow 13.1612\dots \rightarrow \dots$$

Clearly the transition rule from one era to the next is given by:

$$u_{N+1} = \frac{1}{u_N - [u_N]}$$

where $[u_N]$ represents the integer-part of u_N .

How well these discrete maps represent the full nonlinear dynamics has been the subject of much research [10, 11, 12] particularly by S. Rugh and B. Berger.

Writing the BKL parameter in the form $u = [u] + x$ provides a map for the variable x on the unit interval $[0,1]$:

$$x_{n+1} \equiv u_{n+1} - [u_{n+1}] = x_n^{-1} - [x_n^{-1}]$$

which is the well-known Gauss map and its properties have been studied in depth by Khalatnikov et al. [13] and Barrow [14] and Mayer [15]. The Gauss map is interesting in its own right having both periodic behavior as well as chaotic dynamics.

While the discrete maps showed that there was a degree of stochasticity to the Bianchi-IX equations, independent numerical studies of the of the full set of equations arrived at results that were in conflict with each other and with the results obtained from the analysis of the map representation.

The first person to construct numerical solutions to the BKL equations was Zardecki [16] who obtained solutions that showed oscillations in the 3-volume elements. These oscillations were in contradiction to what was expected; i.e. for vacuum solutions the Raychaudhuri equations predicted a monotonic decrease of the 3-volume as

the cosmology approached the singularity. Zardecki's results showed that the primary Lyapunov exponent for the system (a means for measuring chaotic behavior determined by the rate of separation of nearby orbits) was positive which was in agreement with the results obtained from analyzing the Gauss map representation of epoch changes.

The results obtained by Zardecki were often quoted [17, 18] but not called into question until the work of Francisco & Matsas [19], Rugh [10] Burd, et al. [21], and Hobill, et al [22]. Both Rugh and Francisco and Matsas noticed that many of the initial conditions given by Zardecki did not obey the scalar constraint equation. In some cases this could introduce negative energy densities that would act against the collapse to the singularity and set up chaotic oscillations. In addition the calculation by Francisco and Matsas of the dominant Lyapunov exponent showed that it tended toward zero. Similar results were obtained by Burd, et al. and Hobill, et al. showed analytically that the Lyapunov exponents had to be identically zero. The latter also demonstrated that in spite of the fact that constraint obeying initial data was provided in some of Zardecki's simulations, the numerical methods used introduced errors that violated the constraint during the evolution.

The overall conclusion was that a proper numerical treatment of the BKL system in τ time led to vanishing Lyapunov exponents and therefore indicated a lack of chaotic behavior in the system. As pointed out by Pullin [23] and others that the singularity avoiding time coordinate stretched the evolution out such the Kasner epochs became of longer and longer duration thereby killing the chaotic behavior that occurred during the transition from Kasner epoch to Kasner epoch. Thus the use of Lyapunov exponents with different time parametrizations could produce different conclusions about chaotic behavior. The lack of invariance of the time coordinate implies that Lyapunov exponents are not invariant measures of chaotic behavior.

4. HAMILTONIAN DESCRIPTION

Hamiltonian formulations of general relativity have become well established in describing the dynamics associated with the Einstein equations [24]. For Bianchi IX models, the Hamiltonian description has provided a great deal of insight on how the Kasner transitions near the singularity occur.

Misner [25, 26] and others [27] have written the metric (1) in the following manner:

$$ds^2 = -N(t)dt^2 + e^{-2\Omega}(e^{2\beta_{ij}})\omega^i\omega^j \quad (5)$$

where $N(t)$ is the lapse function (often the choice $N = 1$ is used), Ω gives the log of the conformal factor related to the three volume ($e^\Omega = (abc)^{1/3}$) and β_{ij} , the shear matrix is traceless and can be diagonalized such that,:

$$\beta_{ij} = \text{diag}(\beta_+ + \sqrt{3}\beta_-, \beta_+ - \sqrt{3}\beta_-, -2\beta_+,)$$

The relationship between the metric functions a , b , and c and the conformal factor Ω , and shear components is given by:

$$\Omega = -\frac{1}{3} \ln(abc)$$

and

$$\beta_+ = \frac{1}{6} \ln\left(\frac{ab}{c}\right), \quad \beta_- = \frac{1}{2\sqrt{3}} \ln\left(\frac{a}{b}\right)$$

The vacuum field equations for the metric (5) can be written as

$$R_{00} = 3\ddot{\Omega} - 3\dot{\Omega}^2 - 6\dot{\beta}_+^2 - 6\dot{\beta}_-^2 = 0$$

$$\begin{aligned} R_{11} &= -\ddot{\Omega} + \ddot{\beta}_+ + \sqrt{3}\ddot{\beta}_- + 3\dot{\Omega}(\dot{\Omega} - \dot{\beta}_+ - \sqrt{3}\dot{\beta}_-) \\ &+ \frac{1}{2}e^{2\Omega} [e^{4(\beta_+ + \sqrt{3}\beta_-)} - e^{4(\beta_+ - \sqrt{3}\beta_-)} - e^{-8\beta_+} + 2e^{-2(\beta_+ + \sqrt{3}\beta_-)}] = 0 \\ R_{22} &= -\ddot{\Omega} - \ddot{\beta}_+ + \sqrt{3}\ddot{\beta}_- + 3\dot{\Omega}(\dot{\Omega} - \dot{\beta}_+ + \sqrt{3}\dot{\beta}_-) \\ &+ \frac{1}{2}e^{2\Omega} [e^{4(\beta_+ - \sqrt{3}\beta_-)} - e^{4(\beta_+ + \sqrt{3}\beta_-)} - e^{-8\beta_+} + 2e^{-2(\beta_+ - \sqrt{3}\beta_-)}] = 0 \\ R_{33} &= -\ddot{\Omega} - 2\ddot{\beta}_+ + 3\dot{\Omega}(\dot{\Omega} + 2\dot{\beta}_+) \\ &+ \frac{1}{2}e^{2\Omega} [e^{-8\beta_+} - e^{4(\beta_+ + \sqrt{3}\beta_-)} - e^{4(\beta_+ - \sqrt{3}\beta_-)} + 2e^{4\beta_+}] = 0. \end{aligned}$$

Alternatively one can show [27] that the Einstein-Hilbert action ($S = \int R\sqrt{-g}d^4x$) for the diagonal metric (5) in terms of the coordinates β_- and β_+ as well as their conjugate momenta can be written in the form

$$S = \int p_+ d\beta_+ + p_- d\beta_- - H d\Omega$$

where the Hamiltonian, H is given by

$$H = [p_+^2 + p_-^2 + e^{-4\Omega}[V(\beta_+, \beta_-) - 1]]^{1/2} \quad (6)$$

and the “anisotropy potential” V is given by

$$V(\beta_+, \beta_-) = 1 + \frac{1}{3}e^{-8\beta_+} + \frac{2}{3}e^{4\beta_+}[\cosh(4\sqrt{3}\beta_-) - 1] - \frac{4}{3}e^{-2\beta_+} \cosh(2\sqrt{3}\beta_-). \quad (7)$$

The equipotential lines for Eq. (7) plotted on the $[\beta_+, \beta_-]$ space are given in Figure 4 and one also shows how the shear components are related to the a , b and c metric functions introduced in the earlier section.

Using the parameter Ω as a time parameter (i.e often referred to as Omega time) one can use Hamilton’s equations to obtain equations of motion for the variables $(\beta_+, \beta_-, p_+, p_-, H)$

$$d\beta_\pm/d\Omega = \partial H/\partial p_\pm = p_\pm/H$$

$$dp_\pm/d\Omega = -\partial H/\partial \beta_\pm = -\frac{e^{-4\Omega}}{2H} \partial V/\partial \beta_\pm$$

$$dH/d\Omega = -\partial H/\partial \Omega = -2\frac{e^{-4\Omega}}{H}(V - 1)$$

The dynamics of the Hamiltonian system is similar to the “billiard problem” where a hard (elastic) sphere bounces against a closed hardwall container. However in the Bianchi-IX system the triangular potential expands in Ω time and the angle of reflection is not equal to the angle of incidence. As the particle traverses the region

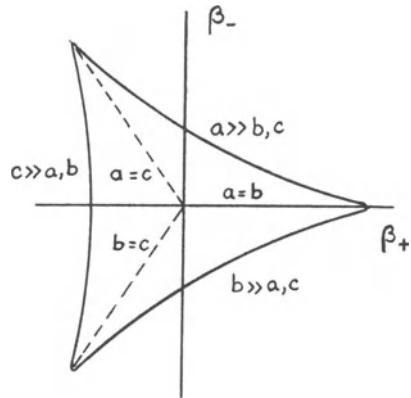


Figure 4. The equipotential lines of the Bianchi IX model in the $(\beta_+ - \beta_-)$ plane. The relationship to the BKL variables is also shown.

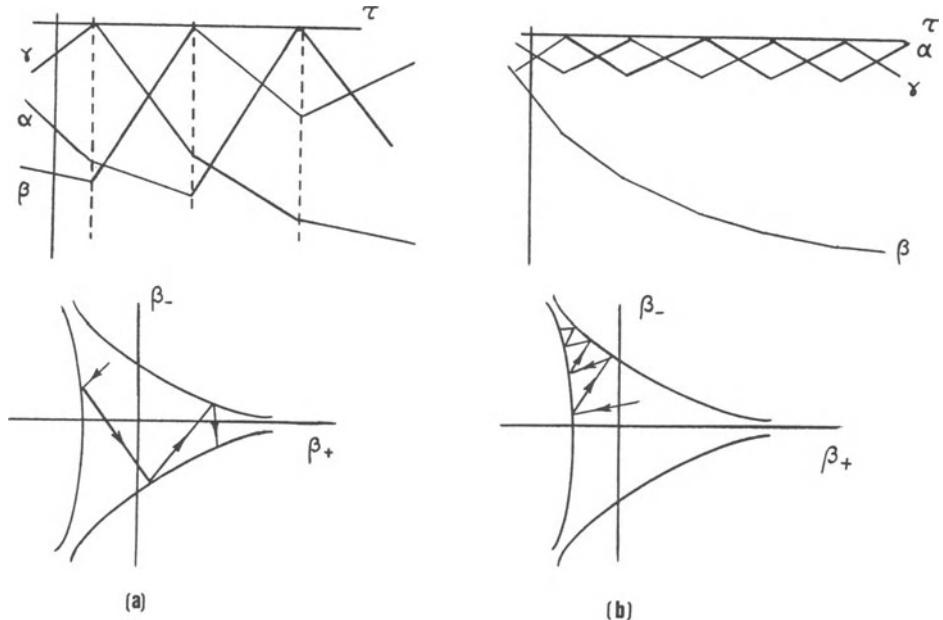


Figure 5. The BKL evolution and the associated bounce maps in the potential walls in the Hamiltonian picture. A generic trajectory that crosses the potential through the middle is shown in (a) while (b) represents bounces that remain in the corner of the potential.

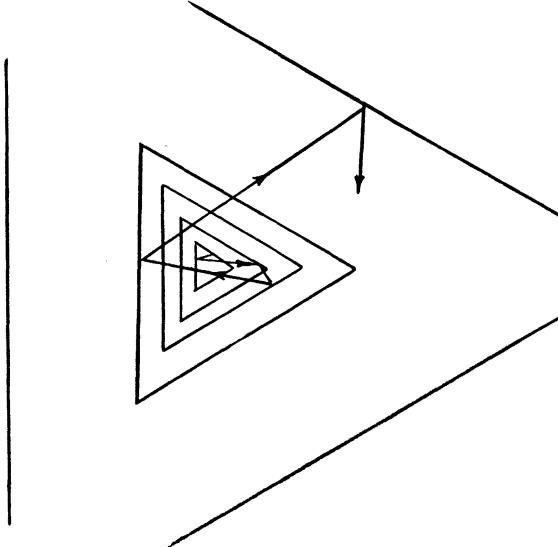


Figure 6. Bounces in an expanding anisotropy potential. The walls of the potential expand outwardly with time and require longer and longer time intervals between wall collisions. As the orbits traverse the potential between the walls, they are approximately in a Kasner epoch.

between the walls where the potential is close to zero and the evolution is approximated by a Kasner solution. The relationship between the Kasner epochs in α , β and γ variables and the bounces at the Mixmaster potential are shown in Figure 5.

If the bounces occur far from the corner channels of the potential, the potential has the approximate form:

$$V \sim \frac{1}{3} e^{-8\beta_+}$$

and one obtains the following discrete bounce law for the angle of reflection θ_r in terms of the angle of incidence, θ_i :

$$\sin \theta_r = 3 \sin \theta_i / (5 - 4 \cos \theta_i)$$

and the change in the Hamiltonian is

$$H_r = \frac{H_i}{3} (5 - 4 \cos \theta_i).$$

The relationship between this approximation and the BKL approximation have been discussed at length by Berger [12]. An example of the trajectory of the particle in the Mixmaster approximate potential is given in Figure 6.

5. THE EMW FORMALISM

Based upon an orthonormal frame method (EMW) introduced by Ellis and MacCallum [28], Wainwright and his collaborators [29, 30, 31] have studied the behavior of the Bianchi IX cosmologies in terms of a different set of ODE and a different discrete map obtained from an approximation of the full equations.

This formalism allows one to link the gravitational degrees of freedom to the shear and expansion tensors (σ_{ik} and θ_{ik} , where $i, k = 1, 2, 3$) of the fluid sources in the spacetime. Adding a scalar constraint equation gives a five-dimensional phase space for the shear variables, Σ_{ik} and the spatial curvature variables N_{ik} where Σ_{ik} and N_{ik} are diagonal and Σ_{ik} is trace-free.

Introducing the spatial curvature as

$$N_{ik} = \text{diag}(N_1, N_2, N_3)$$

and the shear variables Σ_+ and Σ_- defined by:

$$\Sigma_+ = \frac{3}{2}(\Sigma_{22} + \Sigma_{33}); \quad \Sigma_- = \frac{\sqrt{3}}{2}(\Sigma_{22} - \Sigma_{33})$$

The Einstein field equations for this system (dependent upon a time coordinate T defined in terms of the cosmic time t by $dt/dT = 1/H$ where H is the Hubble parameter or e^T is the length scale factor) then become:

$$\begin{aligned}\dot{N}_1 &= (q - 4\Sigma_+)N_1 \\ \dot{N}_2 &= (q + 2\Sigma_+ + 2\sqrt{3}\Sigma_-)N_2 \\ \dot{N}_3 &= (q + 2\Sigma_+ - 2\sqrt{3}\Sigma_-)N_3 \\ \dot{\Sigma}_+ &= -(2 - q)\Sigma_+ - 3S_+ \\ \dot{\Sigma}_- &= -(2 - q)\Sigma_- - 3S_-\end{aligned}$$

where the dot () represents the derivative with respect to T and

$$\begin{aligned}S_+ &= \frac{1}{2}[(N_2 - N_3)^2 - N_1(2N_1 - N_2 - N_3)] \\ S_- &= \frac{\sqrt{3}}{2}[(N_3 - N_2)(N_1 - N_2 - N_3)] \\ q &= \frac{1}{2}(3\gamma - 2)(1 - K) + \frac{3}{2}(2 - \gamma)\Sigma \\ \Sigma &= \Sigma_+^2 + \Sigma_-^2 \\ K &= \frac{3}{4}[N_1^2 + N_2^2 + N_3^2 - 2(N_1N_2 + N_2N_3 + N_3N_1)]\end{aligned}$$

and γ is obtained from a perfect fluid equation of state in the form:

$$p = (\gamma - 1)\rho.$$

These evolution equations have been studied numerically by Ma[32] and Creighton and Hobill [33] and an typical evolution is given in Figure 7.

The system of equations given above have a number of equilibrium points, but the most important is the so-called Kasner-ring which corresponds to the solution a set of Kasner solutions:

$$\Sigma_+^2 + \Sigma_-^2 = 1 \quad N_1 = N_2 = N_3 = 0.$$

Plotted on the Σ_+, Σ_- plane the Kasner circle can be used to define map of the Kasner ring onto itself [34]. The map is determined by the so-called Taub separatrices obtained from the Taub Bianchi II vacuum solutions

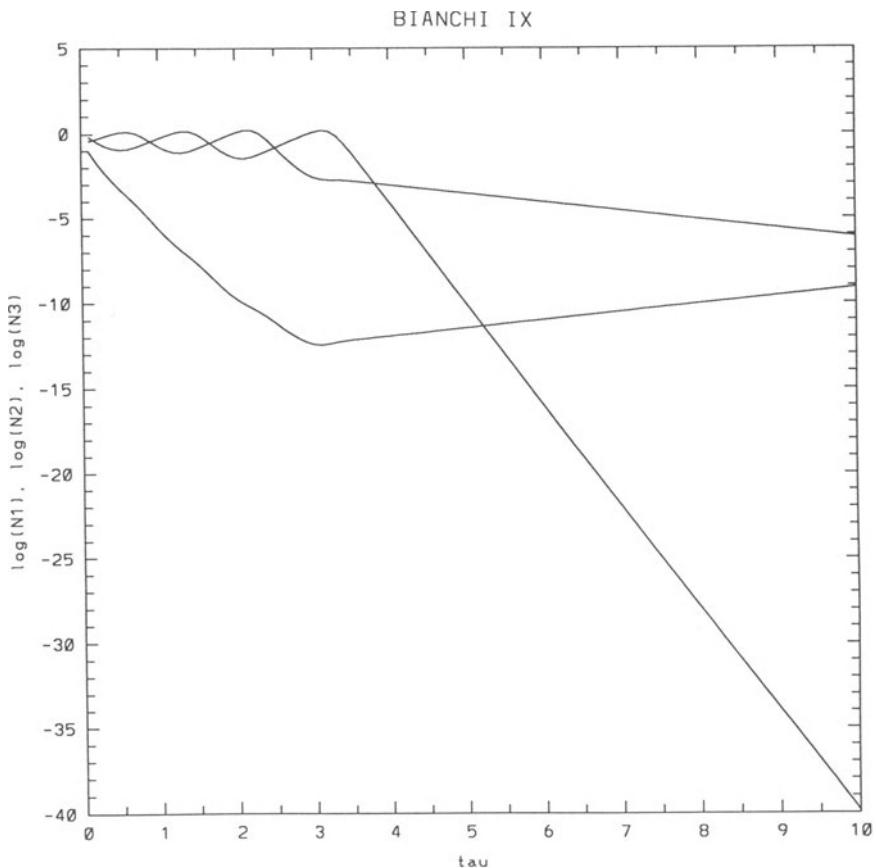


Figure 7. A typical time dependent behavior for the Ellis-MacCallum-Wainwright variables.

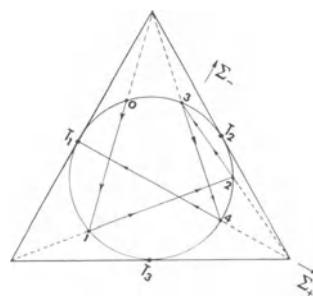


Figure 8. The Bogoyavlenski map (B-map) approximating the EMW evolution. The equilibrium values T_i , ($i = 1, 2, 3$) are associated with special Kasner solutions when the triple (p_1, p_2, p_3) (See Eq. (4)) take on the following values; $T_1 : (1, 0, 0)$, $T_2 : (0, 1, 0)$, and $T_3 : (0, 0, 1)$.

$$\Sigma + K = 1, \quad N_1 > 0 \text{ or } N_1 < 0; \quad N_2 = N_3 = 0$$

and permutations of 1, 2, 3. The projections of the separatrices on the (Σ_+, Σ_-) plane run from the vertices of the equilateral triangle tangent to the Kasner ring to the opposite side of the triangle.

The map that defines the transition from one Kasner epoch to another is determined geometrically in the following manner: A Kasner solution starts on the ring at an arbitrary point. A line is drawn from the ring to the vertex closest to that point and is continued to the opposite side of the triangle. Where the Taub separatrix strikes the Kasner circle a second time defines the (Σ_+, Σ_-) value of the next Kasner epoch. The map then continues on a similar manner: The Figure 8 shows a typical evolution from point $(\Sigma_{+0}, \Sigma_{-0})$ to succeeding points.

How well the map represents the full evolution has been studied by Creighton and Hobill [33]. The exact relationship between the Ellis-MacCallum-Wainwright formalism and the formalisms of BKL and Misner, et al. is still a question open to research.

While this review of the different formulations of the Bianchi-IX equations and the approximations derived from them has been somewhat brief, it is hoped that it will provide some introduction to the notation, and the results obtained over the past few decades of study on these interesting cosmologies.

REFERENCES

- [1] Bombelli, L., and Calzetta, E., (1992), *Class. Quan. Grav.*, **9**, 2573.
- [2] Contopoulos, G., (1990), *Proc. Roy. Soc.*, **431**, 183.
- [3] Choptuik, M., (1993), *Phys. Rev. Lett.*, **70**, 9.
- [4] Calzetta, E., and El-Hasi, C., (1994), *Class. Quan. Grav.*, **10**, 1825.
- [5] MacCallum, M., (1979), in *General Relativity (An Einstein Centenary Survey)*, (Hawking and Israel, eds), CUP, Cambridge, UK.
- [6] Jantzen, R., (1984) in *Cosmology of the Universe*, (Fang and Ruffini, eds), World Scientific, Singapore.
- [7] Belinskii, V.A., Khalatnikov, I.M. and Lifshitz, E.M., (1970), *Adv. Phys.*, **19**, 525.
- [8] Belinskii, V.A., Khalatnikov, I.M. and Lifshitz, E.M., (1971) *Sov. Phys., JETP*, **13**, 745.
- [9] Belinskii, V.A., Khalatnikov, I.M. and Lifshitz, E.M., (1982), *Adv. Phys.*, **31**, 639.
- [10] Rugh, S.E. and Jones, B.J.T., (1990), *Phys. Lett.*, **A**, **147**, 353.
- [11] Berger, B., (1990), *Class. Quan. Grav.*, **7**, 203.
- [12] Berger, B., (1991), *Gen. Rel. Grav.*, **23**, 1385.
- [13] Khalatnikov, I.M., et al., (1985), *J. Stat. Phys.*, **38**, 97.
- [14] Barrow, J.D., (1982), *Phys. Rep.*, **85**, 1.
- [15] Mayer, D.H., (1987), *Phys. Lett.*, **121A**, 390.
- [16] Zardecki, A., (1983) *Phys. Rev.* **D28**, 1235.
- [17] Barrow, J.D., (1984), in *Classical General Relativity*, (Bonnor, Islam & MacCallum, eds.), Cambridge Univ. Press, Cambridge, UK.
- [18] Futamase, T., (1986), *Prog. Theor. Phys.*, **75**, XX.
- [19] Francisco, G. and Matsas, G.E.A., (1988) *Gen. Rel. Grav.* **20** 1047.
- [20] Rugh, S.E., (1989), Cand. Scient. Thesis, Niels Bohr Inst.

- [21] Burd, A., Buric, N., and Ellis, G.F.R., (1990), *Gen. Rel. Grav.*, **22**, 349.
- [22] Hobill, D., Bernstein, D., Simkins, D., and Welge, M., (1991), *Class. & Quan. Grav.* , **8**, 1155.
- [23] Pullin, J., (1991), in *SILARG VII Relativity and Gravitation, Classical and Quantum*, (J. C. D'Olivio, E. Nahmad-Achar, M. Rosenbaum, M. P. Ryan, Jr., L. F. Urrutia and F. Zerrucke, eds.), World Scientific, Singapore, p. 189.
- [24] Arnowitt, R., Deser, S. and Misner, C.W., (1962) in *Gravitation: An Introduction to Current Research*, edited by L. Witten, Wiley, New York.
- [25] Misner, C.W., (1969), *Phys. Rev. Lett.*, **22**, 1071.
- [26] Misner, C.W., (1972) in *Magic without Magic*, (Klauder, ed), Freeman, San Francisco.
- [27] Ryan, M.P. and Shepley, L.C., (1975), *Homogeneous Relativistic Cosmologies*, Princeton Univ. Press, Princeton, NJ.
- [28] Ellis G.F.R. and MacCallum M.A.H., (1969), *Comm. Math. Phys.* **12**, 108.
- [29] Wainwright, J., and Hsu, L.,(1989) *Class. Quan. Grav.*, **6**, 1409.
- [30] Wainwright, J. and Ma, P., (1992) in *Relativity Today*, (ed. Z. Perj'ses), World Scientific, Teaneck, NJ.
- [31] Hewitt, C., and Wainwright, J.,(1993) *Class. Quan. Grav.*, **10**, 99.
- [32] Ma, P.,(1988), M.Sc. Thesis, UWaterloo.
- [33] Creighton, T., & Hobill, D., this volume.
- [34] Bogoyavlensky, O.I., (1985), *Methods in the Qualitative Theory of Dynamical Systems in Astrophysics and Gas Dynamics*, Springer Verlag, New York, NY.

MATHEMATICAL PRELIMINARIES

INTRODUCTION TO DYNAMICAL SYSTEMS¹

John Wainwright

Department of Applied Mathematics, University of Waterloo
Waterloo, Ontario, N2L 3G1 Canada

Abstract. These lecture notes provide an introduction to dynamical systems theory at an advanced undergraduate/graduate level and are intended to serve as a reference for these proceedings. The qualitative behaviour of both linear and non-linear autonomous differential equations is discussed. Particular attention is given to Liapunov stability theory, periodic orbits, limit sets, structural stability, and bifurcation theory, leading up to higher order systems and chaos.

1. INTRODUCTION

We shall review systems of *ordinary differential equations* (DEs) of the form

$$x' = f(x) \tag{1}$$

where $x' = \frac{dx}{dt}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Since the right-hand-side of (1) does not depend on t explicitly, the DE is called *autonomous*. If f is a linear function, i.e.,

$$f(x) = Ax \tag{2}$$

where A is an $n \times n$ matrix of real numbers, the DE is *linear*. In general f will be non-linear. The vector $x \in \mathbb{R}^n$ is called the *state vector* of the system, and \mathbb{R}^n is called the *state space*.

¹Based on Lecture Notes entitled ‘Dynamical Systems’ by J. Wainwright; edited by Alan Coley, Department of Mathematics, Statistics, and Computer Science, Dalhousie University, Halifax, Nova Scotia, B3H 3J5

The function f can be interpreted as a *vector field* on the state space \mathbb{R}^n , since it associates with each $x \in \mathbb{R}^n$ an element $f(x)$ on \mathbb{R}^n , which can be interpreted as a vector

$$f(x) = (f_1(x), \dots, f_n(x)) \quad (3)$$

situated at x .

Definition. A solution of the (DE) (1) is a function $\psi : \mathbb{R} \rightarrow \mathbb{R}^n$ which satisfies

$$\psi'(t) = f(\psi(t)) \quad (4)$$

for all $t \in \mathbb{R}$ (the domain of ψ may be a finite interval (α, β)).

The image of the solution curve ψ in \mathbb{R}^n is called an *orbit* of the (DE). Equation (4) implies that the vector field f at x is tangent to the orbit through x . The state of the physical system that is being analyzed is represented by a point $x \in \mathbb{R}^n$. The evolution of the system in time is described by the motion of this point along an orbit of the DE in \mathbb{R}^n , with t as time. In this interpretation, the DE implies that *the vector field f is the velocity of the moving point in state space* (this should not be confused with the physical velocity of a physical particle).

One cannot hope to find exact solutions of a non-linear DE (1) for $n \geq 2$ (except in very special cases). One thus has to use either qualitative methods, perturbative methods, or numerical methods, in order to deduce the behavior of the physical system. We shall be interested in qualitative methods (in conjunction with ‘numerical experimentation’). The aim of *qualitative analysis* is to understand the qualitative behavior of typical solutions of the DE, for example *the long-term behavior* as $t \rightarrow \infty$ of typical solutions. One is also interested in exceptional solutions such as *equilibrium solutions* or *periodic solutions*, since such solutions can significantly influence the long-term behavior of typical solutions. One is also interested in questions of *stability* and the possible existence of *bifurcations*.

The starting point in the qualitative analysis of an autonomous DE (1) in \mathbb{R}^n is to locate the zeros of the vector field, i.e., to find all $a \in \mathbb{R}^n$ such that

$$f(a) = 0 \quad (5)$$

If $f(a) = 0$, then $\psi(t) = a$, for all $t \in \mathbb{R}$, and it is a solution of the DE, since

$$\psi'(t) = f(\psi(t)) \quad (6)$$

is satisfied trivially for all $t \in \mathbb{R}$. A constant solution $\psi(t) = a$ describes an *equilibrium state* of the physical system, and hence the point $a \in \mathbb{R}^n$ is called an *equilibrium point* of the DE. Here is the official definition.

Definition. Given a DE $x' = f(x)$ in \mathbb{R}^n , any point $a \in \mathbb{R}^n$ which satisfies $f(a) = 0$, is called an *equilibrium point of the DE*.

We are interested in the stability of equilibrium states. In order to address this question it is necessary to study the behaviour of the orbits of the DE close to the equilibrium points. The idea is to consider the linear approximation of the vector field $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ at an equilibrium point. We thus assume that the function f is of class $C^1(\mathbb{R}^n)$ (i.e., that the partial derivatives of f exist and are continuous functions on \mathbb{R}^n .)

Definition. The derivative matrix of $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the $n \times n$ matrix $Df(x)$ defined by

$$Df(x) = \left(\frac{\partial f_i}{\partial x_j} \right), \quad i, j = 1, \dots, n, \quad (7)$$

where the f_i are the component functions of f .

The linear approximation of f is written in terms of the derivative matrix:

$$f(x) = f(a) + Df(a)(x - a) + R_1(x, a), \quad (8)$$

where $Df(a)(x - a)$ denotes the $n \times n$ derivative matrix evaluated at a , acting on the vector $(x - a)$, and $R_1(x, a)$ is the error term. An important result from advanced calculus is that if f is of class C^1 , then the magnitude of the error $\|R_1(x, a)\|$ tends to zero faster than the magnitude of the displacement $\|x - a\|$. Here $\|\cdot\|$ denotes the Euclidean norm on \mathbb{R}^n (i.e., $\|x\| = \sqrt{x_1^2 + \dots + x_n^2}$). This means that in general, $R_1(x, a)$ will be small compared to $Df(a)(x - a)$, for x sufficiently close to a .

If $a \in \mathbb{R}^n$ is an equilibrium point of the DE $x' = f(x)$, we can use (8) to write the DE in the form

$$(NL) : \quad x' = Df(a)(x - a) + R_1(x, a) \quad (9)$$

assuming that f is of class C^1 . We let $u = x - a$, and with the non-linear DE (NL) we associate the linear DE

$$(L) : \quad u' = Df(a)u \quad (10)$$

which is called the linearization of (NL) at the equilibrium point $a \in \mathbb{R}^n$. The question is when do solutions of (L) approximate the solutions of (NL) near $x = a$? i.e., under what conditions can we neglect the error term $R_1(x, a)$? In general the approximation is valid, but in special situations, the approximation can fail. We thus begin with a systematic study of linear DEs.

2. LINEAR AUTONOMOUS DIFFERENTIAL EQUATIONS

The initial value problem in one dimension (i.e., A is a 1×1 matrix or equivalently a constant):

$$x' = Ax, \quad x(0) = a \in \mathbb{R} \quad (11)$$

has the unique solution

$$x(t) = e^{tA}a, \quad \text{all } t \in \mathbb{R}. \quad (12)$$

By analogy, we define the matrix series:

$$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \dots = \sum_{k=0}^{\infty} \frac{1}{k!}A^k \quad (13)$$

called the exponential e^A of A , where A is an $n \times n$ real matrix, I is the $n \times n$ identity matrix and $A^2 = A A$ etc. (matrix product). A matrix series is said to converge if the n^2 infinite series corresponding to the n^2 entries converge in \mathbb{R} . The exponential matrix e^A converges for all $n \times n$ matrices A [cf. Hirsch and Smale, page 83, [1]].

We recall that

$$e^{s+t} = e^s e^t, \quad \text{for all } s, t \in \mathbb{R} \quad (14)$$

(which is proved by using the Taylor series as the definition of e^s , and application of the Binomial Theorem and the Cauchy product for absolutely convergent series). The result (14) does not go over to $n \times n$ matrices due to the general non-commutativity of such matrices. However,

Proposition 1. If A and B are $n \times n$ real matrices, and $AB = BA$, then

$$e^{A+B} = e^A e^B \quad (15)$$

Corollary. If A is an $n \times n$ real matrix, then e^A is invertible and $(e^A)^{-1} = e^{-A}$.

In order to be able to calculate e^A for any matrix, it is necessary to simplify A by performing a similarity transformation:

$$B = P^{-1} A P \quad (16)$$

where P is a non-singular matrix (this corresponds to a change of basis).

Proposition 2. If $B = P^{-1} A P$ then $e^B = P^{-1} e^A P$

Proof. Simplify $(P^{-1} A P)^k$. \square

Proposition 3 (Jordan Canonical Form). For any 2×2 real matrix A , there exists a non-singular matrix P such that

$$J = P^{-1} A P \quad (17)$$

and J is one of the following matrices:

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \quad \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}, \quad (18)$$

Proof. We give an algorithm for constructing P in three mutually exclusive cases, which include all possible 2×2 real matrices; as follows

CASE I: A has two linearly independent eigenvectors f_1, f_2 with eigenvalues λ_1, λ_2 .

Choose $P = (f_1, f_2)$, i.e., f_1, f_2 are the columns of P . Then $AP = (Af_1, Af_2) = (\lambda_1 f_1, \lambda_2 f_2) = P \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. Hence $P^{-1} AP = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$.

CASE II: A has only one eigenvector, with eigenvalue λ . This implies that $(A - \lambda I)^2 = 0$. Choose f_2 such that $(A - \lambda I)f_2 \neq 0$, and let $f_1 = (A - \lambda I)f_2$. Then $\{f_1, f_2\}$ is a basis of \mathbb{R}^2 , and $(A - \lambda I)f_1 = 0$. Choose $P = (f_1, f_2)$. Then $AP = (Af_1, Af_2) = (\lambda f_1, \lambda f_2 + f_1) = P \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$.

CASE III: A admits no real eigenvector, but admits a complex eigenvector: $A(f_1 + if_2) = (\alpha + i\beta)(f_1 + if_2)$, i.e., $Af_1 = \alpha f_1 - \beta f_2$, $Af_2 = \beta f_1 + \alpha f_2$. Choose $P = (f_1, f_2)$. Then $AP = (Af_1, Af_2) = (\alpha f_1 - \beta f_2, \beta f_1 + \alpha f_2) = P \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$. \square

We now have a complete algorithm for calculating e^A for any 2×2 real matrix A :

- (a) find the Jordan canonical form $J = P^{-1} AP$,
- (b) calculate e^J ,
- (c) then $e^A = P e^J P^{-1}$.

2.1. The Flow of a Linear DE

Theorem (Fundamental Theorem for Linear Autonomous DEs). Let A be an $n \times n$ real matrix. Then the initial value problem

$$x' = Ax, \quad x(0) = a \in \mathbb{R}^n \quad (19)$$

has the unique solution

$$x(t) = e^{tA}a, \quad \text{for all } t \in \mathbb{R}. \quad (20)$$

Proof. 1) *Existence:* Let $x(t) = e^{tA}a$ then

$$\frac{dx}{dt} = \frac{d(e^{tA}a)}{dt} = Ae^{tA}a = Ax \quad (21)$$

$$x(0) = e^0a = Ia = a \quad (22)$$

shows that $x(t)$ satisfies the initial value problem (19).

2) *Uniqueness:* Let $x(t)$ be any solution of (19). It follows that

$$\frac{d}{dt} [e^{-tA}x(t)] = 0 \quad (23)$$

Thus $e^{-tA}x(t) = C$, a constant. The initial condition implies that $C = a$ and hence $x(t) = e^{tA}a$. \square

The unique solution of the DE (19) is given by (20) for all t . Thus, for each $t \in \mathbb{R}$, the matrix e^{tA} maps

$$a \rightarrow e^{tA}a \quad (24)$$

(where a is the state at time $t = 0$ and e^{tA} is the state at time t). The set $\{e^{tA}\}_{t \in \mathbb{R}}$ is a 1-parameter family of linear maps of \mathbb{R}^n into \mathbb{R}^n , and is called the *linear flow* of the DE.

We write

$$g^t = e^{tA} \quad (25)$$

to denote the flow. *The flow describes the evolution in time of the physical system for all possible initial states.* As the physical system evolves in time, one can think of the state vector x as a moving point in state space, its motion being determined by the flow $g^t = e^{tA}$. The linear flow satisfies two important properties, which also hold for non-linear flows (to follow).

Proposition 4. The linear flow $g^t = e^{tA}$ satisfies

$$\begin{aligned} F1: \quad g^0 &= I && (\text{identity map}) \\ F2: \quad g^{t_1+t_2} &= g^{t_1} \circ g^{t_2} && (\text{composition}) \end{aligned} \quad (26)$$

Proof. Easy consequence of Proposition 1. \square

Comment: Properties *F1* and *F2* imply that the flow $\{g^t\}_{t \in \mathbb{R}}$ forms a *group* under composition of maps.

The flow g^t of the DE (19) partitions the state space \mathbb{R}^n into subsets called *orbits*, defined by

$$\gamma(a) = \{g^t a | t \in \mathbb{R}\}. \quad (27)$$

The set $\gamma(a)$ is called the *orbit of the DE through a* . It is the image in \mathbb{R}^n of the solution curve $x(t) = e^{tA}a$. It follows that for $a, b \in \mathbb{R}^n$, either $\gamma(a) = \gamma(b)$ or $\gamma(a) \cap \gamma(b) = \emptyset$, since otherwise the uniqueness of solutions would be violated.

Example. Consider $x' = Ax$, $x \in \mathbb{R}^2$, and $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, the linear flow is $e^{tA} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$. The action of the flow on \mathbb{R}^2 , $a \rightarrow e^{tA}a$ corresponds to a clockwise rotation about the origin. Thus if $a \neq 0$, the orbit $\gamma(a)$ is a circle centered at the origin passing through a . The origin is a fixed point of the flow, since $e^{tA}0 = 0$, for all $t \in \mathbb{R}$. The orbit $\gamma(0) = \{0\}$ is called a point orbit. All other orbits are called periodic orbits since $e^{2\pi A}a = a$, i.e., the flow maps onto itself after a time $t = 2\pi$ has elapsed.

Example. Consider $x' = Ax$, $x \in \mathbb{R}^2$, and $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, then the linear flow is $e^{tA} = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$. The action of the flow on \mathbb{R}^2 , $a \rightarrow e^{tA}a$, expands a_1 exponentially and contracts a_2 exponentially, leaving the product a_1a_2 constant. Thus if a is not on one of the axes, the orbit $\gamma(a)$ is a hyperbola. If $a \neq 0$ lies on one of the axes, then the orbit $\gamma(a)$ is a half-axis. The origin is again a point orbit, and all other orbits are non-periodic, i.e., $e^{tA}a \neq a$ for all $t \neq 0$, and $a \neq 0$. Note that the x_1 -axis is the union of three orbits

$$\{(x_1, 0) | x_1 > 0\} \cup \{0\} \cup \{(x_1, 0) | x_1 < 0\}.$$

Classification of orbits of a DE.

1. If $g^t a = a$ for all $t \in \mathbb{R}$, then $\gamma(a) = \{a\}$ and it is called a point orbit. Point orbits correspond to equilibrium points.
2. If there exists a $T > 0$ such that $g^T a = a$, then $\gamma(a)$ is called a periodic orbit. Periodic orbits describe a system that evolves periodically in time.
3. If $g^t a \neq a$ for all $t \neq 0$, then $\gamma(a)$ is called a non-periodic orbit.

Comment:

1. Non-periodic orbits can be of great complexity even for linear DEs if $n > 3$ (for non-linear DEs if $n > 2$).
2. A *solution curve* of a DE is a parameterized curve and hence *contains information about the flow of time t*. The *orbits* are paths in state space (or subsets of state space). Orbits which are not point orbits are *directed paths* with the direction defined by increasing time. The orbits thus do no provide detailed information about the flow of time.

For an autonomous DE, the slope of the solution curves depend only on x and hence the tangent vectors to the solution curves define a vector field $f(x)$ in x -space. Infinitely many solution curves may correspond to a single orbit. On the other hand, a non-autonomous DE does not define a flow or a family of orbits.

2.2. Canonical Linear Flows in \mathbb{R}^2 under Linear Equivalence.

To what extent can a linear DE in \mathbb{R}^n be simplified by making a linear change of

coordinates and a linear change of the time variable? Given a linear DE $x' = Ax$ in \mathbb{R}^n , introduce new coordinates by $y = Px$, where P is a non-singular matrix, and a new time variable $\tau = kt$, where k is a positive constant. It follows that $y' = By$, where $B = \frac{1}{k}PAP^{-1}$.

Definition. The linear DEs $x' = Ax$ and $x' = Bx$ are said to be linearly equivalent if there exists a non-singular matrix P and a positive constant k such that

$$A = kP^{-1}BP. \quad (28)$$

Proposition 5. The linear DEs

$$x' = Ax, \quad x' = Bx, \quad (29)$$

are linearly equivalent if and only if there exists an invertible matrix P and a positive constant k such that

$$P e^{tA} = e^{ktB} P, \quad \text{for all } t \in \mathbb{R}. \quad (30)$$

Proof. Exponentiation of (29) and differentiation of (30). \square

The condition (30), which characterizes linear equivalence, ensures that the linear map P maps each orbit of the flow e^{tA} onto an orbit of the flow e^{tB} .

Definition. Two linear flows e^{tA} and e^{tB} on \mathbb{R}^n are said to be linearly equivalent if there exists a non-singular matrix P and a positive constant k such that

$$P e^{tA} = e^{ktB} P, \quad \text{for all } t \in \mathbb{R}. \quad (31)$$

Let us consider three cases corresponding to the three Jordan canonical forms for any 2×2 real matrix A (see Proposition 3.)

CASE I: TWO EIGENDIRECTIONS. By Proposition 3, there exists a matrix P such that $J = PAP^{-1}$, where

$$J = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

It follows that the given DE is linearly equivalent to $y' = Jy$. The flow is

$$e^{tJ} = \begin{pmatrix} e^{\lambda_1 t} & 0 \\ 0 & e^{\lambda_2 t} \end{pmatrix}.$$

and the eigenvectors are $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. The solutions are $y(t) = e^{tJ}b$, $b \in \mathbb{R}^2$, i.e., $y_1 = e^{\lambda_1 t}b_1$ and $y_2 = e^{\lambda_2 t}b_2$. On eliminating t , we obtain

$$\left(\frac{y_1}{b_1}\right)^{\frac{1}{\lambda_1}} = \left(\frac{y_2}{b_2}\right)^{\frac{1}{\lambda_2}}, \quad \text{if } b_1 b_2 \neq 0 \quad (32)$$

$$y_1 = 0, \quad \text{if } b_1 = 0 \quad (33)$$

$$y_2 = 0, \quad \text{if } b_2 = 0 \quad (34)$$

These equations define the orbits of the DE $y' = Jy$.

- Ia. $\lambda_1 = \lambda_2 < 0$: Attracting Focus (See Fig. 1.)
- Ib. $\lambda_1 < \lambda_2 < 0$: Attracting Node (See Fig. 2.)
- Ic. $\lambda_1 < \lambda_2 = 0$: Attracting Line (See Fig. 3.)
- Id. $\lambda_1 < 0 < \lambda_2$: Saddle (See Fig. 4.)
- Ie. $\lambda_1 = 0 < \lambda_2$: Repelling Line (time reverse of Fig 3.)
- If. $0 < \lambda_1 < \lambda_2$: Repelling Node (time reverse of Fig 2.)
- Ig. $0 < \lambda_1 = \lambda_2$: Repelling Focus (time reverse of Fig 1.)

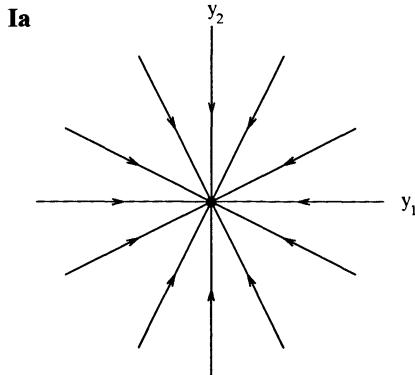


Figure 1. Ia. Attracting Focus

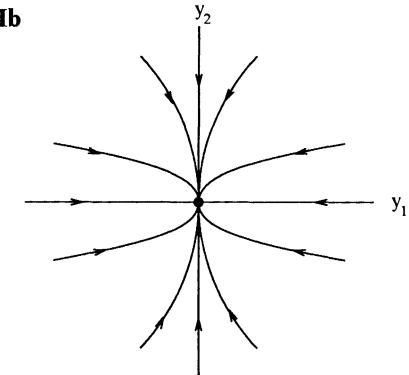


Figure 2. Ib. Attracting Node

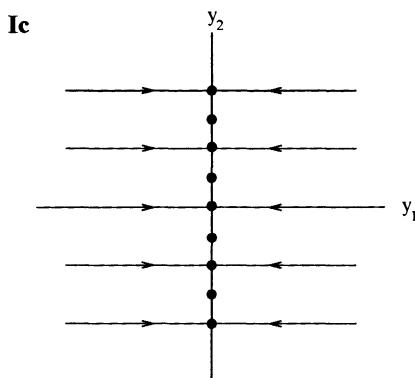


Figure 3. Ic. Attracting Line

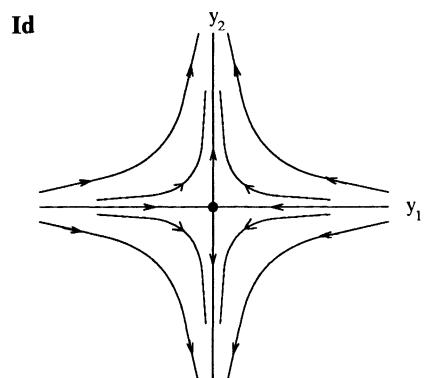


Figure 4. Id. Saddle

CASE II: ONE EIGENDIRECTION. By Proposition 3, there exists a matrix P such that $J = P A P^{-1}$, where

$$J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

It follows that the given DE is linearly equivalent to $y' = Jy$. The flow is

$$e^{tJ} = e^{\lambda t} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

and the single eigenvector is $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We note that if $\lambda \neq 0$, the orbits are given by

$$\begin{aligned} y_1 &= y_2 \left[\frac{b_1}{b_2} + \frac{1}{\lambda} \log \frac{y_2}{y_1} \right], & \text{if } b_2 \neq 0 \\ y_2 &= 0, & \text{if } b_2 = 0 \end{aligned} \quad (35)$$

IIa. $\lambda < 0$: Attracting Jordan Node (See Fig. 5.)

IIb. $\lambda = 0$: Neutral Line (See Fig. 6.)

IIc. $\lambda > 0$: Repelling Jordan Node (time reverse of Fig. 5.)

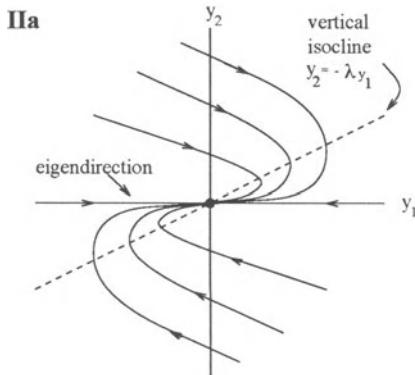


Figure 5. IIa. Attracting Jordan Node

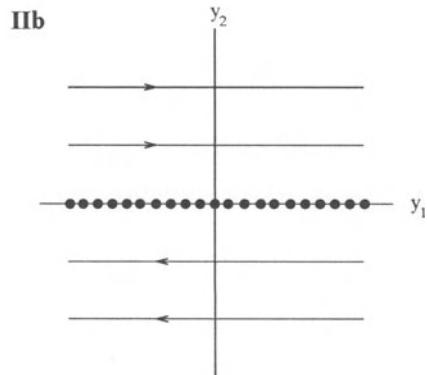


Figure 6. IIb. Neutral Line

CASE III: NO EIGENDIRECTIONS. By Proposition 3, there exists a matrix P such that $J = PAP^{-1}$, where

$$J = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}.$$

It follows that the given DE is linearly equivalent to $y' = Jy$. The simplest way to find the orbits is to introduce polar coordinates (r, θ) : $y_1 = r \cos \theta$, and $y_2 = r \sin \theta$. The DE becomes $r' = \alpha r$ and $\theta' = -\beta$. It follows that $\frac{dr}{d\theta} = -\frac{\alpha}{\beta}r$ which can be integrated to yield $r = r_0 e^{-\frac{\alpha}{\beta}(\theta - \theta_0)}$. Without loss of generality, we can assume $\beta > 0$, since the DE is invariant under the changes $(\beta, y_1) \rightarrow (-\beta, -y_1)$. Thus $\lim_{t \rightarrow \infty} \theta = -\infty$ (counterclockwise rotation as t increases).

IIIa. $\alpha < 0$: Attracting Spiral (See Fig. 7.)

IIIb. $\alpha = 0$: Centre (See Fig. 8.)

IIIc. $\alpha > 0$: Repelling Spiral (time reverse of Fig. 7.)

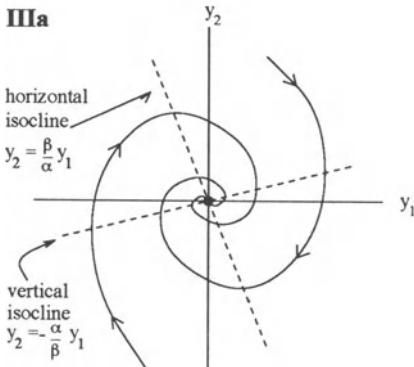


Figure 7. IIIa. Attracting Spiral

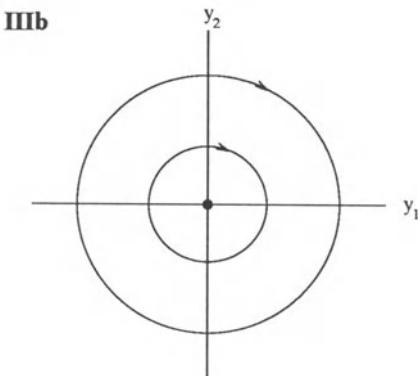


Figure 8. IIIb. Centre

In terms of the Jordan canonical form of two matrices A and B , the corresponding DEs are linearly equivalent if and only if

- A and B have the same number of eigendirections.
- The eigenvalues of A are a multiple (k) of the eigenvalues of B .

This implies that if the DEs have different canonical forms (i.e., belong to different classes Ia–g, IIa–c, IIIa–c) then the DEs are not linearly equivalent. On the other hand, if the DEs have the same canonical form, they will be linearly equivalent if and only if the eigenvalues of A are a multiple of the eigenvalues of B .

Example.

Given DE

$$x' = Ax, \quad A = \begin{pmatrix} -4 & -3 \\ 2 & 1 \end{pmatrix}$$

(See Fig. 9a)

Eigendirections:

$$u_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \quad u_2 = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

Change of variables:

Canonical DE

$$y' = Jy, \quad J = \begin{pmatrix} -2 & 0 \\ 0 & -1 \end{pmatrix}$$

(See Fig. 9b)

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$y = Px, \quad P = \begin{pmatrix} -1 & -1 \\ 2 & 3 \end{pmatrix}, \quad A = P^{-1}JP$$

The flows e^{tA} and e^{tJ} are related by $P e^{tA} = e^{tJ} P$. This implies that the map of the $x_1 x_2$ -plane into the $y_1 y_2$ -plane that is defined by the matrix P , maps orbits of the flow e^{tA} to e^{tJ} .

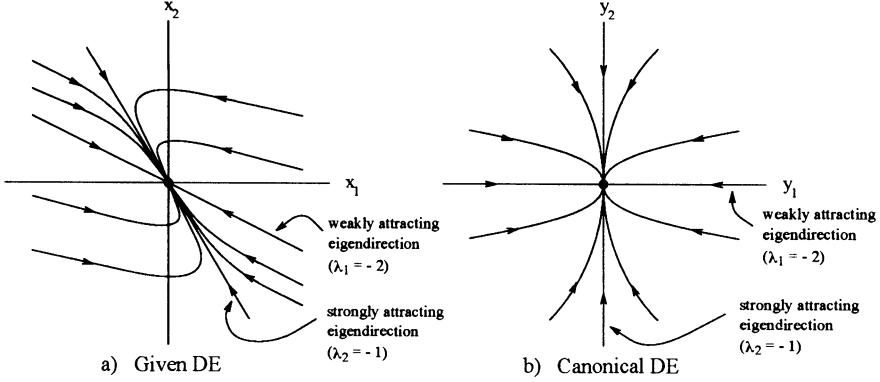


Figure 9. Phase portraits for the Given DE and for the Canonical DE.

2.3. Topological Equivalence

We have seen that under linear equivalence, 2-D flows can be simplified to the extent that they can be parameterized by one real-valued parameter and several discrete parameters (e.g., number of independent eigenvectors). Linear equivalence thus acts as a filter, which retains only certain essential features of the flow i.e., the behavior of the orbits near the equilibrium point $(0, 0)$. On the other hand, if one is primarily interested in long-term behavior, one can use a finer filter, which eliminates more features, and hence leads to a much simpler (but coarser classification). This is the notion of *Topological Equivalence* of linear flows.

For example, cases Ia, Ib, IIa, and IIIa have the common characteristic that all orbits approach the origin (an equilibrium point) as $t \rightarrow \infty$. We would like these flows to be “equivalent” in some sense. We shall show that in fact for all flows of these types, *the orbits of one flow can be mapped onto the orbits of the simplest flow Ia*, using a (non-linear) map $h : \mathbb{R}^2 \rightarrow \mathbb{R}^2$: which is a *homeomorphism on \mathbb{R}^2* .

Definition. A map $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a homeomorphism on \mathbb{R}^n if and only if

1. h is one-to-one and onto,
2. h is continuous,
3. h^{-1} is continuous.

Definition. Two linear flows e^{tA} and e^{tB} on \mathbb{R}^n are said to be topologically equivalent if there exists a homeomorphism h on \mathbb{R}^n and a positive constant k such that

$$h(e^{tA}x) = e^{ktB}h(x), \quad \text{for all } x \in \mathbb{R}^n \text{ and for all } t \in \mathbb{R}. \quad (36)$$

Example. The linear flows e^{tA} , $A = \begin{pmatrix} -2 & 0 \\ 0 & -1 \end{pmatrix}$ and e^{tB} , $B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ are topologically equivalent. The homeomorphism $h : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is given by $y = h(x) =$

$\begin{pmatrix} h_1(x_1) \\ h_2(x_2) \end{pmatrix}$, where $h_1 : \mathbb{R} \rightarrow \mathbb{R}$ and $h_2 : \mathbb{R} \rightarrow \mathbb{R}$ are defined by

$$h_1(x_1) = \begin{cases} \sqrt{x_1}, & \text{if } x_1 \geq 0 \\ -\sqrt{-x_1}, & \text{if } x_1 < 0 \end{cases}, \quad h_2(x_2) = x_2.$$

Definition. A hyperbolic linear flow in \mathbb{R}^2 is one in which the real parts of the eigenvalues are all non-zero (i.e., $\operatorname{Re}(\lambda_i) \neq 0$, $i = 1, 2$).

Proposition 6. Any hyperbolic linear flow in \mathbb{R}^2 is topologically equivalent to the linear flow e^{tA} , where A is one of the following matrices:

$$1. \quad A = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \text{the standard sink.}$$

$$2. \quad A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{the standard source.}$$

$$3. \quad A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{the standard saddle.}$$

Proof. Establish appropriate homeomorphisms. \square

As regards to the *non-hyperbolic* linear flows in \mathbb{R}^2 one can infer by inspection of the portraits that none of the 5 canonical flows [i.e., the centre, the attracting and repelling line, and neutral 2-space ($A = 0$)] are topologically equivalent (their asymptotic behavior as $t \rightarrow \infty$ differs). Thus, two non-hyperbolic linear flows in \mathbb{R}^2 are topologically equivalent if and only if they are linearly equivalent.

Proposition 7. Any non-hyperbolic linear flow in \mathbb{R}^2 is linearly (and hence topologically) equivalent to the flow e^{tA} , where A is one of the following matrices:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (37)$$

These five flows are topologically inequivalent.

2.4. Linear Stability and Linear Sinks in \mathbb{R}^n

Suppose that a physical system in an equilibrium state is disturbed. Does it remain close to (stable) or approach (asymptotically stable) the equilibrium state as time passes ($t \rightarrow \infty$)?

Definition.

1. The equilibrium point 0 of a linear DE $x' = Ax$ in \mathbb{R}^n is stable if for all neighborhoods U of 0 , there exists a neighborhood V of 0 such that $g^t V \subseteq U$ for all $t \geq 0$, where $g^t = e^{tA}$ is the flow of the DE.
2. The equilibrium point 0 of a linear DE $x' = Ax$ in \mathbb{R}^n is asymptotically stable if it is stable and if, in addition, for all $x \in V$, $\lim_{t \rightarrow \infty} \|g^t x\| = 0$.

Proposition 8. Let $A \in M_n(\mathbb{R})$. Then each entry of the matrix e^{tA} is a unique linear combination of the functions

$$t^k e^{\alpha t} \cos(\beta t), \quad t^k e^{\alpha t} \sin(\beta t), \quad (38)$$

where $\alpha + i\beta$ runs through all eigenvalues of A with $\beta \geq 0$ ($\beta = 0$ gives the real eigenvalues) and k takes on all values $0, 1, 2, \dots, n - 1$ less than the multiplicity of the corresponding eigenvalue.

Proof. [cf. Hirsch and Smale, page 135, [1]] \square

Proposition 9 (Characterization of a Sink). Let $A \in M_n(\mathbb{R})$. Then

$$\lim_{t \rightarrow \infty} e^{tA} a = 0 \quad \text{for all } a \in \mathbb{R}^n \quad (39)$$

if and only if $\Re(\lambda) < 0$ for all eigenvalues of A .

Proof. \Rightarrow Suppose that $\Re(\lambda) < 0$, then (39) follows from Proposition 8. \Leftarrow Suppose (39) and that $\lambda = \alpha + i\beta$ is an eigenvalue of A with $\alpha \geq 0$, then we obtain a contradiction using Proposition 8. \square

Equation (39) means that if $\Re(\lambda) < 0$ then *all* solutions $x(t)$ of the DE $x' = Ax$ approach the equilibrium 0 in the long term, i.e.,

$$\lim_{t \rightarrow \infty} x(t) = 0 \in \mathbb{R}^n \quad (40)$$

Thus if $A \in M_n(\mathbb{R})$ is such that $\Re(\lambda) < 0$ for all eigenvalues, then we say that the equilibrium point 0 of the DE $x' = Ax$ is a *sink* in \mathbb{R}^n . If we replace A by $-A$ and t by $-t$, we obtain the time reverse of Proposition 9. Thus if $A \in M_n(\mathbb{R})$ is such that $\Re(\lambda) > 0$ for all eigenvalues, then we say that the equilibrium point 0 of the DE $x' = Ax$ is a *source* in \mathbb{R}^n .

Proposition 10 (Exponential Attraction to a Sink). Let $A \in M_n(\mathbb{R})$. If there exists a constant k such that all eigenvalues of A satisfy $\Re(\lambda) < -k < 0$ then there exists a positive constant M such that

$$\|e^{tA} x\| \leq M e^{-kt} \|x\| \quad \text{for all } x \in \mathbb{R}^n, \quad \text{for all } t \geq 0 \quad (41)$$

Proof. From Proposition 8 and the fact that for any $\epsilon > 0$ and $n > 0$, there exists a constant C such that $t^n < C e^{\epsilon t}$ for all $t \geq 0$. \square

Corollary. If the equilibrium point $0 \in \mathbb{R}^n$ is a sink of the DE $x' = Ax$, then 0 is an asymptotically stable equilibrium point.

Although Proposition 10 guarantees that any initial state is attracted at an exponential rate in time to a linear sink (the equilibrium point 0), it does not imply that the distance from 0, i.e., $\|e^{tA} x\|$ decreases monotonically with t . In other words, as the orbits approach 0, they do not necessarily cut the spheres $\|x\| = R$ in the inward direction. However, as one might expect, one can find a family of concentric ellipsoids, such that as the orbits approach 0, they intersect the ellipsoids in the inward direction.

Proposition 11. If the equilibrium point 0 of the DE $x' = Ax$ in \mathbb{R}^n is a sink, then there exists a positive definite quadratic form

$$V(x) = x^T Q x \quad (42)$$

which is monotone decreasing along all orbits, except for the equilibrium point 0. (Note: Q is a $n \times n$ symmetric matrix such that $V(x) = x^T Q x > 0$ for all $x \neq 0 \in \mathbb{R}^n$. The level sets $V(x) = C > 0$ are ellipsoids in \mathbb{R}^n).

Proof. Differentiate (42) and use Liapunov's Lemma. \square

Let us state Liapunov's Lemma without proof.

Liapunov's Lemma. Let $A \in M_n(\mathbb{R})$. If all eigenvalues of A satisfy $\Re(\lambda) < 0$, then there exists a symmetric positive definite matrix Q such that

$$A^T Q + Q A = -I \quad (43)$$

Comments:

1. The matrix Q in Liapunov's lemma can be found explicitly by solving the linear system of equations (43) for Q .
2. The function V in Proposition 11 is an example of a *Liapunov function* for the equilibrium point 0. Such functions will play an important role when we discuss non-linear stability later.
3. Proposition 11 can also be used to prove that any linear sink in \mathbb{R}^n is topologically equivalent to the standard sink.

3. NON-LINEAR DIFFERENTIAL EQUATIONS

For non-linear DEs, one does not expect to be able to write down the flow explicitly. Indeed the aim of the subject of dynamical systems is to describe the qualitative properties of a non-linear flow *without knowing the flow explicitly*.

Example. We shall illustrate a quick way to draw the orbits for a 1-D DE. Consider the 1-D DE $x' = f(x) = x(1-x)$, it is the sign of $f(x)$ that determines the direction of each orbit which corresponds to increasing t . (See Fig. 10)

Example. Consider the non-linear DE in \mathbb{R}^2 , $x' = f(x)$ where

$$f(x) = \begin{pmatrix} x_1(1-x_1) \\ -2x_2 \end{pmatrix} \quad \text{and } x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

or equivalently

$$x'_1 = x_1(1-x_1), \quad x'_2 = -2x_2.$$

The 2-D flow $\{g^t\}$ is defined by

$$g^t a = \begin{pmatrix} \frac{e^t a_1}{e^t a_1 + 1} \\ e^{-2t} a_2 \end{pmatrix}, \quad \text{where } a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \quad (44)$$

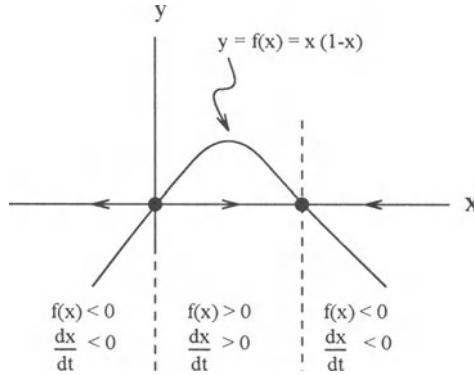


Figure 10. Quick way to draw orbits for a 1-D DE.

$\{g^t\}$ is a one-parameter family of non-linear maps of $\mathbb{R}^2 \rightarrow \mathbb{R}^2$. Since the set of values of t for which $\{g^t a\}$ is defined depends on a , $\{g^t\}$ is called the local flow. We note that the flow satisfies the same group properties as in the linear case. Because the DEs are un-coupled, one can also obtain an explicit expression for the orbits in \mathbb{R}^2 . By eliminating t we obtain

$$\frac{dx_1}{dx_2} = -\frac{x_1(1-x_1)}{2x_2}, \quad \text{for } x_2 \neq 0,$$

and hence

$$x_1^2 x_2 = k(1-x_1)^2, \quad k = \text{constant}. \quad (45)$$

There are two equilibrium points, i.e., $f(x) = 0$:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

These are necessarily fixed points of the flow:

$$g^t \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad g^t \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

as may be verified using (44). By studying the flow (44), and equation (45) for the orbits, one can obtain a qualitative sketch of the orbits (see Fig. 11). Note that the behavior of the orbits near the equilibrium points can be inferred by approximating equation (45):

if $x_1 \approx 0$, then $x_1^2 x_2 \approx k$

if $x_1 \approx 1$, then $x_2 \approx k(1-x_1)^2$

Note that if the initial state a satisfies $a_1 > 0$, then

$$\lim_{t \rightarrow \infty} g^t a = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

i.e., the long term behaviour of the system is to approach the equilibrium state $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

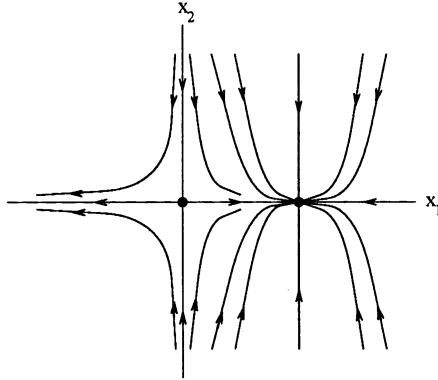


Figure 11. Phase portrait for the system $x'_1 = x_1(1 - x_1)$, $x'_2 = -2x_2$.

3.1. Hamiltonian DEs in 2-D.

We now discuss a class of DEs in \mathbb{R}^2 whose orbits can be studied directly, even though the flow cannot be found explicitly. Consider a particle moving in 1-D under the influence of a force which depends only on position. If q denotes the position of the particle, and m its mass, then Newton's Second Law states that

$$m \frac{d^2q}{dt^2} = F(q). \quad (46)$$

Since $F = F(q)$, by assumption, there exists a function $V(q)$ defined up to an additive constant, called the *potential function*, such that

$$F(q) = -V'(q). \quad (47)$$

In general, we introduce the *linear momentum*

$$p = m \frac{dq}{dt}. \quad (48)$$

Then the equation of motion (46), with (47), leads to the 2-D DE

$$q' = \frac{1}{m}p, \quad p' = -V'(q). \quad (49)$$

The state vector of the system (mass moving subject to the force) is $(q, p) \in \mathbb{R}^2$. The energy of the system is the sum of kinetic and potential energy:

$$\frac{1}{2}m \left(\frac{dq}{dt} \right)^2 + V(q). \quad (50)$$

When expressed in terms of q and p , the energy is referred to as the *Hamiltonian of the system*, denoted by H :

$$H(q, p) = \frac{1}{2m}p^2 + V(q). \quad (51)$$

In terms of H , the DE (49) assumes the form

$$q' = \frac{\partial H}{\partial p}, \quad p' = -\frac{\partial H}{\partial q}. \quad (52)$$

These equations are called *Hamilton's equations* for the mechanical system. An important consequence of (52) is that H is constant along any solution curve:

$$\frac{dH}{dt} = \frac{\partial H}{\partial q} \frac{dq}{dt} + \frac{\partial H}{\partial p} \frac{dp}{dt} = \frac{\partial H}{\partial q} \frac{\partial H}{\partial p} + \frac{\partial H}{\partial p} \left(-\frac{\partial H}{\partial q} \right) = 0. \quad (53)$$

This expresses *conservation of energy*, and is expected in view of the physical situation.

Let us make some comments on the orbits of a Hamiltonian DE (52), with H of the form (51),

- Since H is constant along any solution curve, the orbits in q - p -space are contained in the level sets of the Hamiltonian function, given by $H(q, p) = C$.
- The *equilibrium points* of the DE are given by $p = 0$ and $V'(q) = 0$, i.e., they are determined by the critical points of $V(q)$.
- The isocline corresponding to $q' = 0$ (vertical tangent lines) is given by $p = 0$, and the isocline(s) corresponding to $p' = 0$ (horizontal tangent lines) are given by $V'(q) = 0$.
- For orbits with $H(q, p) = C$, the q -values are determined by the inequality $V(q) \leq C$, as follows from equation (51).

Example: A magneto-elastic beam. [cf. Guckenheimer and Holmes, page 83 [2]]
The Hamiltonian is

$$H(q, p) = \frac{1}{2m}p^2 - \frac{1}{2}\beta^2q^2 + \frac{1}{4}q^4,$$

and the Hamiltonian DE (the equations of motion) is

$$q' = \frac{1}{m}p, \quad p' = \beta^2q - q^3.$$

where q denotes the displacement of the beam from line of symmetry and the force gives rise to the potential $V(q) = -\frac{1}{2}\beta^2q^2 + \frac{1}{4}q^4$. The phase portrait is given by Fig. 12.

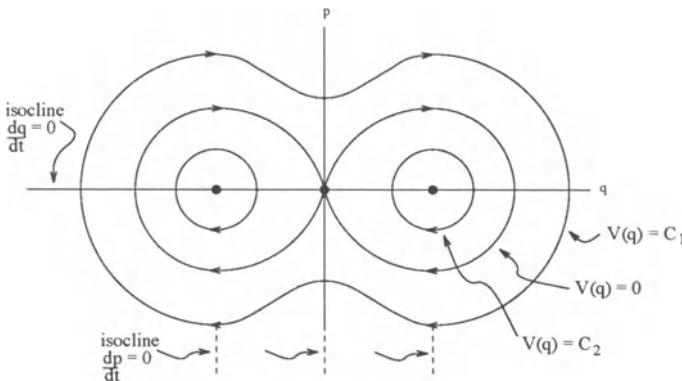


Figure 12. Phase portrait for the magneto-elastic beam.

Terminology: Any 2-D DE of the form (52) is called a *Hamiltonian DE*, irrespective of whether the Hamiltonian $H(q, p)$ is of the form (51). If H is of the form (51), the

resulting DE describes a mechanical system, and will be called a *classical Hamiltonian DE*. The concept of a Hamiltonian DE can be generalized to higher dimensions:

$$q'_i = \frac{\partial H}{\partial p_i}, \quad p'_i = -\frac{\partial H}{\partial q_i}, \quad (54)$$

where the state vector is $x = (q_1, \dots, q_n, p_1, \dots, p_n) \in \mathbb{R}^{2n}$. The motion of a spherical pendulum, or motion of a particle in a plane under a central potential (e.g., planetary motion) leads to a Hamiltonian DE with $n = 2$ i.e., state space is \mathbb{R}^4 .

3.2. The Flow of a Non-Linear DE

We begin by stating the standard existence-uniqueness theorem for the initial value problem (IVP) for a DE in \mathbb{R}^n .

Theorem (Existence-Uniqueness). Consider the IVP

$$x' = f(x), \quad x(0) = a \in \mathbb{R}^n. \quad (55)$$

If $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is of class $C^1(\mathbb{R}^n)$, then for all $a \in \mathbb{R}^n$, there exists an interval $(-\delta, \delta)$ and a unique function $\psi_a : (-\delta, \delta) \rightarrow \mathbb{R}^n$ such that

$$\psi'_a(t) = f(\psi_a(t)), \quad \psi_a(0) = a. \quad (56)$$

Proof. The idea is to rewrite the IVP as an integral equation, and use Picard iterates. [cf. Hirsch and Smale, page 162, [1]] \square

Comment: If the hypothesis (f be of class C^1) is weakened, then uniqueness may fail, (e.g., the DE $x' = x^{\frac{2}{3}}$ in \mathbb{R} , has two solutions which satisfy the initial condition $x(0) = 0$, namely $x(t) = 0$ and $x(t) = \frac{1}{27}t^3$. Note that $f(x) = x^{\frac{2}{3}}$ is continuous but not C^1 .

The existence-uniqueness theorem is a local result — it guarantees existence of a solution in some interval $(-\delta, \delta)$ centered at $t = 0$. Since we are interested in the long-term behaviour of solutions, we would like the solutions to be defined for all $t \geq 0$. We can extend the interval of definition of the solution $\psi_a(t)$ by successively reapplying the theorem, and in this way obtain a *maximal interval of definition* of the solution $\psi_a(t)$. We shall denote this maximal interval by (α, β) .

Definition. We say that the solution $\psi_a(t)$ has finite escape time β_a if

$$\lim_{t \rightarrow \beta_a^-} \|\psi_a(t)\| = +\infty \quad (57)$$

Theorem (Maximality). Let $\psi_a(t)$ be the unique solution of the DE $x' = f(x)$, where $f \in C^1(\mathbb{R}^n)$, which satisfies, $\psi_a(0) = a$, and let (α_a, β_a) denote the maximal interval on which $\psi_a(t)$ is defined. If β_a is finite, then

$$\lim_{t \rightarrow \beta_a^-} \|\psi_a(t)\| = +\infty \quad (58)$$

Proof. [cf. Hirsch and Smale, pages 171–172, [1]] \square

Corollary. Consider the DE $x' = f(x)$, $f \in C^1(\mathbb{R}^n)$. If a solution $\psi_a(t)$ is bounded for $t \geq 0$, then the solution is defined for all $t \geq 0$.

Comment: One can always modify a given DE $x' = f(x)$, $x \in \mathbb{R}^n$, and $f \in C^1(\mathbb{R}^n)$, so that the orbits are unchanged, but such that all solutions are defined for all $t \in \mathbb{R}$. The idea is to re-scale the vector field f (the velocity of the state point x):

$$f(x) \rightarrow \lambda(x)f(x), \quad (59)$$

where $\lambda(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is a C^1 -function (a scalar) which is *positive* on \mathbb{R}^n (in order to preserve the direction of time). This rescaling does not change the direction of the vector field, hence the orbits are unchanged. However, one can choose λ so that $\|\lambda f\|$ is bounded e.g.,

$$\lambda(x) = \frac{1}{1 + \|f(x)\|}. \quad (60)$$

Proposition 12. If $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is $C^1(\mathbb{R}^n)$, and $\lambda : \mathbb{R}^n \rightarrow \mathbb{R}$ is $C^1(\mathbb{R}^n)$ and positive, then $x' = f(x)$ and $x' = \lambda(x)f(x)$ have the same orbits, and λ can be chosen so that all solutions of the second DE are defined for all $t \in \mathbb{R}$.

Proof. [cf. Nemytskii and Stepanov, Theorem 3.22, page 19, and Theorem 1.31, page 9, [3]] \square

Definition. Consider a DE $x' = f(x)$, where f is of class $C^1(\mathbb{R}^n)$, whose solutions are defined for all $t \in \mathbb{R}$. Let $\psi_a(t)$ be the unique maximal solution which satisfies $\psi_a(0) = a$. The flow of the DE is defined to be the one-parameter family of maps $\{g^t\}_{t \in \mathbb{R}}$ such that $g^t : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and $g^t a = \psi_a(t)$ for all $a \in \mathbb{R}^n$.

The flow $\{g^t\}$ is defined in terms of the solution function $\psi_a(t)$ of the DE by

$$g^t a = \psi_a(t). \quad (61)$$

It is important to understand the difference between $\psi_a(t)$ and $g^t a$ conceptually:

- For a fixed $a \in \mathbb{R}^n$, $\psi_a : \mathbb{R}^n \rightarrow \mathbb{R}^n$ gives the state of the system $\psi_a(t)$ for all $t \in \mathbb{R}$, with $\psi_a(0) = a$ initially.
- For a fixed $t \in \mathbb{R}$, $g^t : \mathbb{R}^n \rightarrow \mathbb{R}^n$ gives the state of the system $g^t a$ at time t for all initial states a .

The solution function $\psi_a(t)$ satisfies $\psi'_a(t) = f(\psi_a(t))$, $\psi_a(0) = a$. Hence $\psi'_a(0) = f(a)$. By definition of the flow, it follows that

$$\left. \frac{d}{dt}(g^t a) \right|_{t=0} = f(a), \quad (62)$$

which is simply a statement of the fact that the vector field f is tangent to the orbits of the DE.

Proposition 13. Let $\{g^t\}$ be the flow of a DE $x' = f(x)$, then

$$\begin{aligned} F1: \quad g^0 &= I && (\text{identity map}) \\ F2: \quad g^{t_1+t_2} &= g^{t_1} \circ g^{t_2} && (\text{composition}) \end{aligned} \quad (63)$$

Proof. Use the translational property of solutions of an autonomous DE. \square

Theorem (Smoothness of a Flow). If $f \in C^1(\mathbb{R}^n)$, then the flow $\{g^t\}$ of the DE $x' = f(x)$ consists of C^1 maps.

Proof. [cf. Hirsch and Smale, pages 298–300, [1]] □

Comment: The significance of this result is that the solutions of the DE depend smoothly on the initial state.

Definition. The orbit through a , denoted $\gamma(a)$ is defined to be

$$\gamma(a) = \{x \in \mathbb{R}^n | x = g^t a, \text{ for all } t \in \mathbb{R}\} \quad (64)$$

As in the linear case, orbits are classified as *point orbits*, *periodic orbits*, and *non-periodic orbits*. Sometimes it is convenient to work with the *positive orbit through a* denoted $\gamma^+(a)$ and defined by

$$\gamma^+(a) = \{x \in \mathbb{R}^n | x = g^t a, \text{ for all } t \geq 0\} \quad (65)$$

3.3. Long-Term Behaviour and Limit Sets

Consider a physical system with initial state vector $x \in \mathbb{R}^n$, whose evolution is described by a DE $x' = f(x)$, which determines a flow $\{g^t\}_{t \in \mathbb{R}}$. A fundamental question is: What is the *long-term behaviour* of the system as $t \rightarrow \infty$, starting at an initial state a when $t = 0$? In other words, what happens to the positive orbit through a defined by (65) as $t \rightarrow +\infty$?

The simplest behaviour is that the system, starting at state a , *approaches an equilibrium state* as $t \rightarrow \infty$, i.e., $\lim_{t \rightarrow \infty} g^t a = p$. In this case, we say that the ω -limit set of the initial point a is the equilibrium point p , and write

$$\omega(a) = \{p\} \quad (66)$$

The next simplest behaviour is that the system, starting at state a , approaches periodic evolution, i.e., the orbit approaches a periodic orbit γ . In this situation, $\lim_{t \rightarrow \infty} g^t a$ does not exist, since the orbit does not approach a unique point. However, for any point $p \in \gamma$, we can choose a sequence of times $\{t_n\}$, with $\lim_{n \rightarrow \infty} t_n = \infty$, such that $\lim_{n \rightarrow \infty} g^{t_n} a = p$. In this case we say that the ω -limit set of the initial point a is the periodic orbit γ , and write

$$\omega(a) = \gamma. \quad (67)$$

These examples motivate the definition to follow.

Definition. Consider the DE $x' = f(x)$ in \mathbb{R}^n , and the associated flow $\{g^t\}_{t \in \mathbb{R}}$. Given an initial point $a \in \mathbb{R}^n$, a point $p \in \mathbb{R}^n$ is said to be an ω -limit point of a if there exists a sequence $\{t_n\}$ with $\lim_{n \rightarrow \infty} t_n = \infty$ such that $\lim_{n \rightarrow \infty} g^{t_n} a = p$. The set of all ω -points of a is called the ω -limit set of a , denoted by $\omega(a)$.

Example. Consider the system

$$x'_1 = x_1(1 - x_1), \quad x'_2 = -x_2.$$

The equilibrium points are $(0, 0)$ and $(1, 0)$. For $a = (a_1, a_2)$ we have

$$\begin{aligned} a_1 > 0, \quad \omega(a) &= (1, 0); \\ a_1 = 0, \quad \omega(a) &= (0, 0); \\ a_1 < 0, \quad \omega(a) &= \emptyset. \end{aligned}$$

To help in identifying the ω -limit set of an initial state a , we consider the following question: What subsets of \mathbb{R}^n can be ω -limit sets for a flow $\{g^t\}$? This is a difficult question, and is unsolved if $n > 2$. But there is a simple necessary condition which is indispensable in identifying $\omega(a)$.

Proposition 14. An ω -limit set $\omega(a)$ of a flow $\{g^t\}$ is a whole orbit of the flow, or is the union of more than one whole orbit.

Proof. We simply prove that if $y \in \omega(a)$, then the orbit through y given by

$$\gamma(y) = \{g^t y \mid t \in \mathbb{R}\} \quad (68)$$

is contained in $\omega(a)$. \square

It is also important to know that an ω -limit set is non-trivial (i.e., not the empty set).

Proposition 15. If the positive orbit through a ,

$$\gamma^+(a) = \{g^t a \mid t \geq 0\} \quad (69)$$

is bounded, then $\omega(a) \neq \emptyset$.

Proof. By the Bolzano-Weierstrass theorem, the bounded set $\{g^n a \mid n \in \mathbb{N}\}$ has at least one limit point. \square

3.4. Trapping Sets and the Global Liapunov Theorem

In this section we discuss a method for locating the ω -limit sets of a certain class of DEs, namely those which admit a so-called *Liapunov function*.

Definition. Given a DE $x' = f(x)$ in \mathbb{R}^n , a set $S \subseteq \mathbb{R}^n$ which is the union of whole orbits of the DE, is called an invariant set for the DE.

For example if we have a Hamiltonian DE in \mathbb{R}^2 , then the level sets $H(x_1, x_2) = k$ are invariant sets, since H is constant along any orbit. More generally, we have the concept of a *first integral*.

Definition. A function $H : \mathbb{R}^n \rightarrow \mathbb{R}$ of class C^1 , that is not constant on any open subset of \mathbb{R}^n , is called a first integral of the DE $x' = f(x)$ if H is constant on every orbit:

$$\frac{d}{dt} H(x(t)) = 0 \quad \text{for all } t. \quad (70)$$

Since

$$\frac{d}{dt} H(x(t)) = \nabla H(x(t)) \cdot f(x(t)), \quad (71)$$

using the chain rule and the DE, it follows that $H(x)$ is a first integral of the DE $x' = f(x)$ if and only if

$$\nabla H(x) \cdot f(x) = 0 \quad \text{for all } x \in \mathbb{R}^n, \quad (72)$$

and $H(x)$ is not identically constant on any open subset of \mathbb{R}^n . If one has a first integral (e.g., a Hamiltonian function) then the orbits of the DE are contained in the one-parameter family of level sets $H(x) = k$.

It sometimes happens that one has a function $F : \mathbb{R}^n \rightarrow \mathbb{R}$ such that *only a particular level set of F is an invariant set*. For example look at the DE, $x'_1 = x_1(1 - x_1)$, $x_2 = -2x_2$, and consider the functions $F(x_1, x_2) = x_1$ and $G(x_1, x_2) = x_2$. The level sets $F = 0$ and $F = 1$ are invariant sets since $x_1 = 0$ and $x_1 = 1$ satisfy the DE; but F is not a first integral. Similarly the level set $G = 0$ is an invariant set. These invariant sets play a major role in determining the portrait of the orbits.

Proposition 15. Given a DE $x' = f(x)$, in \mathbb{R}^n , and a function $G : \mathbb{R}^n \rightarrow \mathbb{R}$ of class C^1 . If $\nabla G(x) \cdot f(x) = 0$ for all x such that $G(x) = k$, then the level set $G(x) = k$ is an invariant set of the DE.

Proof. The vector field f is tangent to the level set $G(x) = k$, and hence for any initial state a with $G(a) = k$, the orbit $\gamma(a)$ lies in the level set. \square

In order to determine an ω -limit set, it is helpful to know that an orbit enters a bounded set S and *never leaves it*. Such a set is called a trapping set.

Definition. Given a DE $x' = f(x)$ in \mathbb{R}^n , with flow g^t , a subset $S \subset \mathbb{R}^n$ is said to be a trapping set of the DE if it satisfies

1. S is a closed and bounded set,
2. $a \in S$ implies $g^t a \in S$ for all $t \geq 0$.

The usefulness of trapping sets lies in this result; if S is a trapping set of a DE $x' = f(x)$, then for all $a \in S$, the ω -limit set $\omega(a)$ is non-empty and is contained in S .

Example. Consider the DE

$$x'_1 = \gamma_1(1 - x_1 - \alpha x_2)x_1, \quad x'_2 = \gamma_2(1 - \beta x_1 - x_2)x_2,$$

with $x_1 \geq 0$, $x_2 \geq 0$, $\gamma_1, \gamma_2, \alpha$ and β are positive constants with $\alpha < 1$ and $\beta < 1$. By inspection we see that the x_1 -axis ($x_2 = 0$) and the x_2 -axis ($x_1 = 0$) are invariant sets. By inspection, for sufficiently large x_1 and/or x_2 , then $x'_1 < 0$ and $x'_2 < 0$. Thus the set

$$S_k = \{(x_1, x_2) | x_1 + x_2 \leq k, x_1 \geq 0, x_2 \geq 0\}$$

is a trapping set for the DE.

3.5. The Global Liapunov Theorem

Consider a DE $x' = f(x)$ in \mathbb{R}^n and let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be $C^1(\mathbb{R}^n)$. We can calculate the rate of change of V along a solution of the DE:

$$\begin{aligned} \frac{d}{dt} V(x(t)) &= \frac{\partial V}{\partial x_1} \frac{dx_1}{dt} + \cdots + \frac{\partial V}{\partial x_n} \frac{dx_n}{dt} \quad \text{by the Chain Rule} \\ &= \nabla V(x(t)) \cdot f(x(t)) \equiv \dot{V}(x) \end{aligned} \quad (73)$$

using $x'_i = f_i$ and the definition of scalar product in \mathbb{R}^n . Suppose that $\dot{V}(x) \leq 0$ for all $x \in \mathbb{R}^n$. Then for any orbit $\gamma(a)$ in a trapping set S , $V(x)$ will keep decreasing along $\gamma(a)$ until the orbit approaches its ω -limit set $\omega(a)$. One thus expects that $\omega(a)$ will consist of points for which $\dot{V}(x) = 0$. In this way, one obtains a strong restriction on the possible ω -limit sets.

Theorem (Global Liapunov Theorem). Consider the DE $x' = f(x)$ in \mathbb{R}^n , and let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a C^1 function. If $S \subset \mathbb{R}^n$ is a trapping set, and $\dot{V}(x) \leq 0$ for all $x \in S$, then for all $a \in S$, $\omega(a) \subseteq \{x \in S | \dot{V}(x) = 0\}$.

Proof. [cf. Hale, Theorem 1.3, page 296, [4]] □

Comments:

1. A function $V : \mathbb{R}^n \rightarrow \mathbb{R}$ which satisfies the above theorem for $x \in S \subset \mathbb{R}^n$ is called a *Liapunov function on S* .
2. One can often use the level sets of the function V to define the trapping set S in the theorem.
3. In applying the theorem, we note that we simply have to find *whole orbits* that are contained in the set $\{x \in S | \dot{V}(x) = 0\}$ to obtain the ω -limit set $\omega(a)$.

Example. Consider the DE

$$x'_1 = x_2, \quad x'_2 = -\alpha x_2 - x_1^3, \quad \alpha > 0.$$

We note that when $\alpha = 0$, the DE is Hamiltonian, with $H = \frac{1}{2}x_2^2 + \frac{1}{4}x_1^4$. So we let $V(x) = \frac{1}{2}x_2^2 + \frac{1}{4}x_1^4$. It follows that $\dot{V}(x) = -\alpha x_2^2 \leq 0$ on \mathbb{R}^2 . The level sets $V(x) = k$ are simple closed curves and $\nabla V(x)$ points outwards. Thus $S_k = \{x \in \mathbb{R}^2 | V(x) \leq k\}$ is a trapping set. Thus for all $a \in S_k$, $\omega(a) \subseteq \{x \in S_k | \dot{V}(x) = 0\} = \{x \in S_k | x_2 = 0\}$. However, when $x_2 = 0$, the DE implies that $x'_2 \neq 0$ unless $x_1 = 0$. Thus the equilibrium point $(0, 0)$ is the only whole orbit with $x_2 = 0$, and hence $\omega(a) = \{(0, 0)\}$, for all $a \in S_k$. Finally, note that for all $a \in \mathbb{R}^2$, $a \in S_k$ for some k . Thus $\omega(a) = \{(0, 0)\}$, for all $x \in \mathbb{R}^2$.

4. LIAPUNOV'S STABILITY THEOREM

The goal is to show that the stability of an equilibrium point can be ascertained, subject to a condition, by studying the linearization of the DE. The basic definitions are the same as in the linear case, with the linear flow e^{tA} being replaced by g^t .

Definition.

1. The equilibrium point \bar{x} of a DE $x' = f(x)$ in \mathbb{R}^n is stable if for all neighborhoods U of \bar{x} , there exists a neighborhood V of \bar{x} such that $g^t V \subseteq U$ for all $t \geq 0$, where g^t is the flow of the DE.
2. The equilibrium point \bar{x} of a DE $x' = f(x)$ in \mathbb{R}^n is asymptotically stable if it is stable and if, in addition, for all $x \in V$, $\lim_{t \rightarrow \infty} \|g^t x - \bar{x}\| = 0$

Now, consider a non-linear DE $x' = f(x)$ in \mathbb{R}^n . Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a $C^1(\mathbb{R}^n)$. The rate of change of V along a solution of the DE is given by

$$\frac{d}{dt}V(x(t)) = \nabla V(x(t)) \cdot f(x(t)) \equiv \dot{V}(x) \quad (74)$$

Thus, if $\dot{V}(x) < 0$ for all t then $V(x)$ decreases with time along the corresponding orbit. From a geometrical point of view, the orbits cut the level sets $V(x) = k$ in the direction away from $\nabla V(x)$. Suppose that \bar{x} is an equilibrium point of the DE. If $V(\bar{x}) = 0$ and $V(x) > 0$ for all $x \in U - \{\bar{x}\}$, where U is a neighborhood of \bar{x} , then we expect the level sets of V in U to be concentric curves ($n=2$) or concentric spheres ($n=3$); consequently when $\dot{V} < 0$ for all $x \in U - \{\bar{x}\}$, any orbit in $U - \{\bar{x}\}$ will cut the level sets of V in the inward direction, and we expect that this will continue until the orbit is forced to approach the equilibrium point \bar{x} as $t \rightarrow \infty$, showing that the equilibrium point is *asymptotically stable*. If, instead, $\dot{V} \leq 0$ for all $x \in U - \{\bar{x}\}$, then U may contain periodic orbits, and we only obtain the weaker conclusion that \bar{x} is *stable*. Finally, if $\dot{V} > 0$ for all $x \in U - \{\bar{x}\}$, then the orbits are forced away from \bar{x} , which is thus an *unstable* equilibrium point.

Theorem (Liapunov Stability Theorem). *Let \bar{x} be an equilibrium point of the DE $x' = f(x)$ in \mathbb{R}^n . Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a C^1 function such that $V(\bar{x}) = 0$, $V(x) > 0$ for all $x \in U - \{\bar{x}\}$, where U is a neighborhood of \bar{x} .*

1. *If $\dot{V}(x) < 0$ for all $x \in U - \{\bar{x}\}$, then \bar{x} is asymptotically stable.*
2. *If $\dot{V} \leq 0$ for all $x \in U - \{\bar{x}\}$, then \bar{x} is stable.*
3. *If $\dot{V}(x) > 0$ for all $x \in U - \{\bar{x}\}$, then \bar{x} is unstable.*

Proof. This can be proved as a corollary of the Global Liapunov Theorem. \square

A function $V : \mathbb{R}^n \rightarrow \mathbb{R}$ which satisfies $V(\bar{x}) = 0$, $V(x) > 0$ for all $x \in U - \{\bar{x}\}$, and $\dot{V}(x) \leq 0$ (respectively < 0) for all $x \in U - \{\bar{x}\}$, is called a *Liapunov function* (respectively, a *strict Liapunov function*) for the equilibrium point \bar{x} . Hence we obtain the following:

Theorem (Criterion for Asymptotic Stability). *Let \bar{x} be an equilibrium point of the DE $x' = f(x)$ in \mathbb{R}^n . If all eigenvalues of the derivative matrix $Df(\bar{x})$ satisfy $\Re(\lambda) < 0$, then the equilibrium point \bar{x} is asymptotically stable.*

Proof. Consider the linear approximation of $f(x)$ at \bar{x} , $A(x - \bar{x})$, where $A = Df(\bar{x})$. Since $\Re(\lambda) < 0$ for all eigenvalues of A , by Liapunov's Lemma there exists a symmetric positive definite matrix Q such that $A^T Q + Q A = -I$. To complete the proof, we note that $V(x) = x^T Q x$ is a strict Liapunov function for \bar{x} . \square

4.1. Linearization and the Hartman-Grobman Theorem

Consider the DE, $x'_1 = x_1(1 - x_1)$, $x'_2 = -2x_2$, studied earlier. We consider the orbits of the linearizations at the equilibrium points $(0, 0)$ and $(1, 0)$. At $(0, 0)$, $A = Df(0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$, a saddle. At $(1, 0)$, $A = Df(1, 0) = \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}$, an attracting node. For each equilibrium point, there is a homeomorphism h which maps the orbits

of the linearized flow in a neighborhood of O onto the orbits of the non-linear flow in a neighborhood of the equilibrium point. In other words, the linearizations give a reliable description of the non-linear orbits *near the equilibrium points*.

Let $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a C^2 function. Suppose that \bar{x} is a critical point of g , i.e., $\nabla g(\bar{x}) = 0$. Let

$$Hg(\bar{x}) = \begin{pmatrix} g_{11}(\bar{x}) & g_{12}(\bar{x}) \\ g_{21}(\bar{x}) & g_{22}(\bar{x}) \end{pmatrix} \quad (75)$$

be the Hessian matrix of g at \bar{x} . The second derivative test determines whether \bar{x} is a local maximum (respectively local minimum, saddle point) of g , subject to a certain restriction, namely $\det[Hg(\bar{x})] \neq 0$, where the second derivative test may fail. In a similar manner, the linearization of a non-linear DE can fail to give reliable information about the orbits, if a certain restriction does not hold.

Example. Consider the non-linear DE $x'_1 = -x_1$, $x'_2 = x_2^3$, which describes a non-linear saddle with orbits $x_1 = ke^{\frac{1}{2}x_2^2}$. The linearization at $(0,0)$ is $u' = Au$, $A = Df(0,0) = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$, which describes an attracting line.

The linear and non-linear flows are not topologically equivalent in a neighborhood of the equilibrium point, and hence the linearization fails. The source of the failure is that the matrix $Df(0,0)$ has a zero eigenvalue.

Example. Consider the DE $x'_1 = -x_2 - x_1(x_1^2 + x_2^2)$, $x'_2 = x_1 - x_2(x_1^2 + x_2^2)$, a non-linear spiral. The linearization at $(0,0)$ is $u' = Au$, $A = Df(0,0) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, representing a centre.

Again, the linear and non-linear flows are not topologically equivalent in a neighborhood of the equilibrium points, hence the linearization fails. The source of the failure is that the matrix $Df(0,0)$ has eigenvalues with zero real parts.

Theorem (Hartman-Grobman). Let \bar{x} be an equilibrium point of the DE $x' = f(x)$ in \mathbb{R}^n , where $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is of class C^1 . If all the eigenvalues of the matrix $Df(\bar{x})$ satisfy $\Re(\lambda) \neq 0$, then there is a homeomorphism $h : U \rightarrow \bar{U}$ of a neighborhood U of O onto a neighborhood \bar{U} of \bar{x} which maps orbits of the linear flow $e^{tDf(\bar{x})}$ onto orbits of the non-linear flow g^t of the DE, preserving the parameter t .

Proof. [cf. Hartman, pages 244-250, [5]] □

The Hartman-Grobman Theorem can be stated more concisely using the concept of topological equivalence, which can be generalized to non-linear flows.

Definition. Two flows g^t and \tilde{g}^t on \mathbb{R}^n are said to be topologically equivalent if there is a homeomorphism $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$ which maps orbits of g^t onto orbits of \tilde{g}^t , and preserves the direction of the parameter t .

Then we can state: If \bar{x} is a *hyperbolic equilibrium point*, then the flow of the DE $x' = f(x)$ and the flow of its linearization $u' = Df(\bar{x})u$, are *locally* topologically equivalent.

Comment: An equilibrium point \bar{x} of a non-linear DE is said to be *hyperbolic* if all eigenvalues of the matrix $Df(\bar{x})$ satisfy $\Re(\lambda) \neq 0$.

4.2. Saddle Points and the Stable Manifold Theorem

Definition. An equilibrium point \bar{x} of a DE $x' = f(x)$ in \mathbb{R}^n is a saddle point if the real parts of the eigenvalues of the matrix $Df(\bar{x})$ are all non-zero, and not all of one sign. [i.e., a saddle point is a hyperbolic (all $\Re(\lambda) \neq 0$) equilibrium point which is neither a sink (all $\Re(\lambda) < 0$) nor a source (all $\Re(\lambda) > 0$).]

The Hartman-Grobman theorem gives a qualitative local description of a (non-linear) saddle. In particular in \mathbb{R}^2 we have (see Fig. 13).

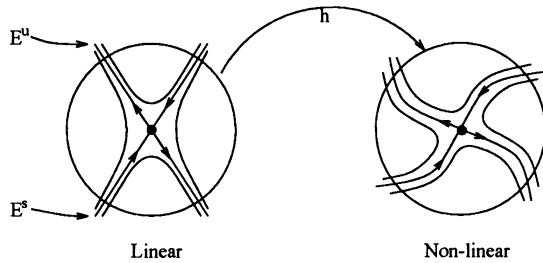


Figure 13. The linear and non-linear flows in a neighborhood of a saddle point.

The 1-D subspace E^s that is spanned by the eigenvector which corresponds to the eigenvalue $\lambda_1 < 0$ is called the *stable subspace* of the equilibrium point, and the 1-D subspace E^u that is spanned by the eigenvector which corresponds to the eigenvalue $\lambda_2 > 0$ is called the *unstable subspace*.

Definition. Let \bar{x} be a saddle point of the DE $x' = f(x)$ in \mathbb{R}^n , and let U be a neighborhood of \bar{x} . The local stable manifold of \bar{x} in U is defined by

$$W^s(\bar{x}, U) = \left\{ x \in U \mid g^t x \xrightarrow[t \rightarrow \infty]{\text{ }} \bar{x}, g^t x \in U \text{ for all } t \geq 0 \right\}. \quad (76)$$

We can now state the following theorem (without proof).

Theorem (Stable Manifold Theorem). Let \bar{x} be a saddle point of $x' = f(x)$ in \mathbb{R}^n , where f is of class C^1 , and let E^s be the stable subspace of the linearization at \bar{x} . Then there exists a neighborhood U of \bar{x} such that the local stable manifold $W^s(\bar{x}, U)$ is a smooth (C^1) curve which is tangent to E^s at \bar{x} .

Comment: One can define, in an analogous way the local unstable manifold of \bar{x} in U denoted $W^u(\bar{x}, U)$, and similarly there is an “Unstable Manifold Theorem.”

4.3. Local Behaviour Near a Non-linear Sink

Suppose that \bar{x} is an equilibrium point of a non-linear DE $x' = f(x)$ in \mathbb{R}^2 . Suppose that all eigenvalues of the matrix $Df(\bar{x})$ satisfy $\Re(\lambda) < 0$, (i.e., \bar{x} is a sink). The Hartman-Grobman theorem asserts that in some neighborhood of \bar{x} , the flow of the non-linear DE is topologically equivalent to the flow of the linearization $u' = Df(\bar{x})u$, where $u = x - \bar{x}$. In this section we give a more detailed description of the non-linear orbits near a sink.

Let $\bar{x} = (\bar{x}_1, \bar{x}_2)$ be an asymptotically stable equilibrium point of the DE $x' = f(x)$. In order to describe the orbits near \bar{x} , we introduce polar coordinates

$$\begin{aligned} x_1 - \bar{x}_1 &= r \cos \theta \\ x_2 - \bar{x}_2 &= r \sin \theta \end{aligned} \quad (77)$$

Since \bar{x} is asymptotically stable,

$$\lim_{t \rightarrow +\infty} r(t) = 0 \quad (78)$$

if $r(0)$ is sufficiently close to zero. We say that the equilibrium point \bar{x} is a *non-linear spiral* if

$$\lim_{t \rightarrow +\infty} \theta(t) = \pm\infty \quad (79)$$

for any solution $(r(t), \theta(t))$ for which (78) holds.

Proposition 16. Consider the DE

$$(NL) : \quad x' = f(x) \quad (80)$$

in \mathbb{R}^2 , where f is of class C^1 . Consider the linearization

$$(L) : \quad u' = Df(\bar{x})u \quad (81)$$

at the equilibrium point \bar{x} . If O is an attracting spiral point of (L) , then \bar{x} is an attracting spiral point of (NL) .

Proof. [cf. Coddington and Levinson, Theorem 2.2, page 376 [6]] □

Proposition 17. If O is an attracting node of (L) then \bar{x} is an attracting node of (NL) .

Proof. [cf. Coddington and Levinson, Theorem 5.1, page 384, [6]] □

Comment: A similar result holds for Jordan nodes. [cf. Coddington and Levinson, page 387, [6]]

An asymptotically stable equilibrium point \bar{x} is said to be an *attracting non-linear focus* if all orbits sufficiently close to \bar{x} approach \bar{x} in a definite direction as $t \rightarrow \infty$, and given any direction there exists an orbit which tends to \bar{x} in this direction.

Comment: If O is a focus of (L) , it does not necessarily follow in general that \bar{x} is a non-linear focus of (NL) .

Proposition 18. Suppose that the vector field f is of class C^2 . If O is an attracting focus of (L) then \bar{x} is an attracting focus of (NL) .

Proof. [cf. Coddington and Levinson, page 377, [6]] □

A stable equilibrium point \bar{x} is said to be a *non-linear centre* if in some neighborhood of \bar{x} , the orbits are periodic orbits which enclose \bar{x} . Recall that the Hartman-Grobman theorem does not apply if O is a *centre* of (L) , i.e., one cannot conclude that \bar{x} is a *centre* of (NL) . But one can still draw a useful conclusion.

Proposition 19. If O is a centre of (L) , then \bar{x} is either a centre, an attracting spiral, or a repelling spiral of (NL) .

Proof. [cf. Coddington and Levinson, Theorem 4.1, page 382, [6]] □

5. PERIODIC ORBITS AND LIMIT SETS IN THE PLANE

We have seen that a linear DE can admit a family of periodic orbits, corresponding to a physical system whose motions are undamped oscillations. Of greater interest is the case where a DE admits an *isolated periodic orbit*, i.e., the orbit has a neighborhood U which contains no other periodic orbits. This is in fact only possible for a non-linear DE. In this situation, the periodic orbit γ may attract neighboring orbits, thereby describing a physical system which has an *oscillatory steady state* which is stable. We say that such a system undergoes *self-sustained oscillations*. The main goal in this section is to discuss isolated periodic orbits of DEs in the plane.

It should be noted that the question of existence of periodic orbits is a difficult one. In 1900, as part of problem 16 of his famous list, David Hilbert posed the question: What is the maximum number of *isolated* periodic orbits of an autonomous DE $x' = f(x)$ in \mathbb{R}^2 , if the components of the vector field f are *polynomial functions*? This problem is still unsolved even for the case of quadratic polynomials (degree 2). For awhile the upper bound was thought to be 3 but an example with 4 isolated periodic orbits has been found.

5.1. Non-Existence of Periodic Orbits

Dulac's criterion for excluding periodic orbits for a DE in \mathbb{R}^2 is based on Green's theorem.

Theorem (Green's). If g_1 and g_2 are of class C^1 on an open set $D \subset \mathbb{R}^2$ and C is a simple closed curve in D , whose interior R is in D then

$$\oint_C g_1 dx_1 + g_2 dx_2 = \iint_R \left(\frac{\partial g_2}{\partial x_1} - \frac{\partial g_1}{\partial x_2} \right) dx_1 dx_2 \quad (82)$$

where C is oriented counter-clockwise.

Recall that the line integral is evaluated as an ordinary integral by introducing parametric equations for C :

$$\oint_C g_1 dx_1 + g_2 dx_2 = \int_{t=a}^b [g_1(x_1(t), x_2(t))x'_1(t) + g_2(x_1(t), x_2(t))x'_2(t)] dt \quad (83)$$

where $(x_1(b), x_2(b)) = (x_1(a), x_2(a))$. Since (x'_1, x'_2) is tangent to C it follows that if the vector field (g_1, g_2) is orthogonal to C at each point of C then

$$\oint_C g_1 dx_1 + g_2 dx_2 = 0 \quad (84)$$

The idea is to apply Green's theorem to a periodic orbit γ of the DE $x' = f(x)$ in \mathbb{R}^2 , where $f(x) = (f_1(x), f_2(x))$. The essential point is to note that the vector field $(f_2(x), -f_1(x))$ is orthogonal to the periodic orbit γ , since (f_1, f_2) is tangent to γ , and $(f_1, f_2) \cdot (f_2, -f_1) = 0$. We apply Green's theorem to a periodic orbit γ , with $(g_1, g_2) = (f_2, -f_1)$. It follows that

$$\iint_R \left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} \right) dx_1 dx_2 = 0 \quad (85)$$

Thus, if the DE is such that $\text{div}(f) = \nabla \cdot f = \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} > 0 (< 0)$, for all $x \in D$, where D is a simply connected open set, then a contradiction arises, and we conclude that D contains no periodic orbits. The requirement that D be simply connected is necessary in order to ensure that the interior of γ is contained in D .

The preceding argument may be generalized by noting that for any scalar function B , the vector field $(Bf_2, -Bf_1)$ is orthogonal to the periodic orbit γ . The condition which excludes periodic orbits becomes $\text{div}(Bf) = \nabla \cdot Bf = \frac{\partial Bf_1}{\partial x_1} + \frac{\partial Bf_2}{\partial x_2} > 0 (< 0)$.

Proposition 20 (Dulac's Criterion). *If $D \subseteq \mathbb{R}^2$ is a simply connected open set and $\text{div}(Bf) = \frac{\partial}{\partial x_1}(Bf_1) + \frac{\partial}{\partial x_2}(Bf_2) > 0, (< 0)$ for all $x \in D$ where B is a C^1 function, then the DE $x' = f(x)$ where $f \in C^1$ has no periodic orbit which is contained in D .*

Proof. Essentially given above. □

Comment: The function $B(x_1, x_2)$ is called a *Dulac function* for the DE in the set D .

Example. A classical Hamiltonian DE in 2-D,

$$x'_1 = x_2, \quad x'_2 = -V'(x_1)$$

typically admits a family of periodic orbits. Modify the DE by adding linear damping:

$$x'_1 = x_2, \quad x'_2 = -\alpha x_2 - V'(x_1) \quad \alpha < 0.$$

It follows that

$$\text{div}(f) = \frac{\partial}{\partial x_1}(x_2) + \frac{\partial}{\partial x_2}(-\alpha x_2 - V'(x_1)) = -\alpha < 0$$

for all $x \in \mathbb{R}^2$. Thus the damped DE admits no periodic orbits, irrespective of the form of the potential $V(x_1)$.

The second criterion for excluding periodic orbits, which is valid in \mathbb{R}^n , $n \geq 2$, follows from the observation that if a function $V(x)$ is monotone decreasing along an orbit of a DE, then that orbit cannot be periodic.

Proposition 21. *Let $V : \mathbb{R}^n \rightarrow \mathbb{R}$ be a C^1 function. If $\dot{V}(x) = \nabla V(x) \cdot f(x) \leq 0$ on a subset $D \subseteq \mathbb{R}^n$, then any periodic orbit of the DE $x' = f(x)$ which lies in D , belongs to the subset $\{x | \dot{V}(x) = 0\} \cap D$.*

5.2. Self-Sustained Oscillations–An Example

We consider the DE

$$x'' + (x^2 - 1)x' + x = 0 \quad \text{or} \quad \begin{cases} x'_1 = x_2, \\ x'_2 = -x_1 - (x_1^2 - 1)x_2. \end{cases}$$

This DE, the famous *Van der Pol DE*, arises in the study of electrical circuits. It admits an isolated periodic orbit which is not easy to locate. So we begin by considering a modification of the DE, for which the periodic orbit can be found explicitly.

Consider the DE, $x'_1 = x_2$, $x'_2 = -x_1 - (x_1^2 + x_2^2 - 1)x_2$. the origin $(0, 0)$ is the only equilibrium point. In order to determine whether orbits approach $(0, 0)$ we consider $V(x_1, x_2) = x_1^2 + x_2^2$. Along any solution

$$\frac{d}{dt}V(x_1(t), x_2(t)) = -[V(x_1(t), x_2(t)) - 1]x_2^2. \quad (86)$$

$V(x_1(t), x_2(t)) = 1$ satisfies this DE; and hence the circle $x_1^2 + x_2^2 = 1$ is a periodic orbit. Further any annulus $1 \leq x_1^2 + x_2^2 \leq R$ is positively invariant, and V is a Liapunov function i.e., $\dot{V} \leq 0$ by (86). It follows from the Global Liapunov theorem that for any initial state a exterior to the circle $x_1^2 + x_2^2 = 1$, the ω -limit set $\omega(a)$ is the periodic orbit $x_1^2 + x_2^2 = 1$. Similarly, the annulus $0 < x_1^2 + x_2^2 \leq 1$ is positively invariant, and it follows from equation (86) that V is a Liapunov function. Thus, for any initial state $a \neq 0$ inside the circle $x_1^2 + x_2^2 = 1$, the ω -limit set $\omega(a)$ is the periodic orbit $x_1^2 + x_2^2 = 1$. This type of periodic orbit is called a limit cycle.

Definition. An isolated periodic orbit γ of a DE $x' = f(x)$ in \mathbb{R}^2 , is called a stable limit cycle if there exists a neighborhood U of γ such that $\omega(a) = \gamma$ for all $a \in U$.

5.3. The Poincaré-Bendixson Theorem

Consider the DE $x'_1 = x_2$, $x'_2 = -x_1 + (1 - x_1^2 - \frac{1}{4}x_2^2)x_2$. Let $V(x_1, x_2) = x_1^2 + x_2^2$. It follows that $\dot{V}(x_1, x_2) = (1 - x_1^2 - \frac{1}{4}x_2^2)x_2^2$. On the circle $C_1 : x_1^2 + x_2^2 = 1$, it follows that $\dot{V}(x_1, x_2) = \frac{3}{4}x_2^4 \geq 0$. On the circle $C_2 : x_1^2 + x_2^2 = 4$ we have that $\dot{V}(x_1, x_2) = -\frac{3}{4}x_1^2x_2^2 \leq 0$. It follows that the annulus S bounded by C_1 and C_2 is a trapping set. Thus for any initial state $a \in S$, the ω -limit set $\omega(a)$ must be a whole orbit (or the union of whole orbits), we conjecture that $\omega(a)$ must be a periodic orbit. The validity of this conjecture constitutes the Poincaré-Bendixson Theorem.

Before discussing the theorem, we need a few preliminary concepts.

Definition. A local section of the flow of a DE in \mathbb{R}^2 is a smooth curve segment Σ such that the vector field f of the DE satisfies $n \cdot f \neq 0$ on Σ , where n is normal to Σ .

Comment: This implies that no equilibrium points of f lie on Σ , and by continuity, that orbits pass through Σ in one direction only.

Definition. Let x be a regular (i.e., non-equilibrium) point of the flow i.e., $f(x) \neq 0$. Let Σ be a local section through x . A flow-box for x is a neighborhood of x of the form $N = \left\{ g^t \Sigma \mid |t| < \delta \right\}$ for some $\delta > 0$.

Finally we need the following properties of ω -limit sets:

1. $\omega(a)$ is the union of whole orbits,
2. $\omega(a)$ is a closed set,
3. if $\omega(a)$ is bounded, then $\omega(a)$ is connected (i.e., is not the union of disjoint sets),
4. the following Lemma.

Lemma (Fundamental Lemma on ω -limit sets in \mathbb{R}^2). Let $\omega(a)$ be an ω -limit set of a DE in \mathbb{R}^2 . If $y \in \omega(a)$, then the orbit through y , $\gamma(y)$, cuts any local section Σ in at most one point.

Proof. [cf. Hirsch and Smale, Proposition 2, page 246, [1]] also [cf. Hale, Corollary 1.1, page 53, [4]] \square

Comment: The Lemma is not valid for a flow in \mathbb{R}^3 (a local section is a smooth surface segment, for a flow on a 2-torus).

Theorem (Poincaré-Bendixson). Let $\omega(a)$ be a non-empty ω -limit set of the DE $x' = f(x)$ in \mathbb{R}^2 , where $f \in C^1$. If $\omega(a)$ is a bounded subset of \mathbb{R}^2 and $\omega(a)$ contains no equilibrium points, then $\omega(a)$ is a periodic orbit.

Proof. [cf. Hirsch and Smale, Chapter 11, [1]] \square

In applications it is often convenient to use the following Corollary of the Poincaré-Bendixson theorem.

Corollary. Let K be a positively invariant subset of the DE $x' = f(x)$ in \mathbb{R}^2 , where $f \in C^1$. If K is a closed and bounded set, then K contains either a periodic orbit or an equilibrium point.

There is one further result which can help to locate isolated periodic orbits.

Proposition 21. Any periodic orbit of a C^1 DE on \mathbb{R}^2 encloses an equilibrium point.

5.4. The Van der Pol DE and Liénard's Theorem

In 1922, a Dutch scientist, Balthasar van der Pol, published a paper concerning oscillations in radio circuits containing triode valves (now obsolete). The analysis was based on the DE

$$x'' + \mu(x^2 - 1)x' + x = 0, \quad \mu > 0 \quad (87)$$

now known as the *van der Pol DE*, and used as a simple model for systems which can undergo self-sustained oscillations. The French scientist Alfred Liénard, was also interested in self-sustained oscillations and in 1928 published an analysis of a DE with a more general dissipative term:

$$x'' + g(x)x' + x = 0, \quad (88)$$

now known as *Liénard's DE*.

If we let $G(x) = \int_0^x g(s) ds$ then the Liénard DE is equivalent to

$$x'_1 = x_2 - G(x_1), \quad x'_2 = -x_1 \quad (89)$$

where, in the case of the van der Pol DE, $G_p(x) = \mu(\frac{1}{3}x^3 - x)$. The existence of an isolated periodic orbit of (89) depends on the shape of the vertical isocline, which is given by $x_2 = G(x_1)$. The requirement is that this curve should be qualitatively the same as the van der Pol isocline, i.e., $x_2 = G_p(x_1)$ for large x_1 and for x_1 close to 0.

Following Liénard, consider the following conditions:

L1: G is an odd function,

L2: $\lim_{x \rightarrow \infty} G(x) = +\infty$, and there exists $\beta > 0$ such that $G(x) > 0$, $G'(x) > 0$ for $x > \beta$.

L3: There exists an α , with $0 < \alpha \leq \beta$, such that $G(x) < 0$ for $0 < x < \alpha$.

The corollary to the Poincaré-Bendixson theorem can then be used to prove the existence of a periodic orbit of the DE (89).

Theorem (Liénard). *If G satisfies conditions L1–L3, then the DE (88) admits a periodic orbit.*

Proof. [cf. Hale, pages 57-59, [4]] □

Comment: If $\alpha = \beta$, in conditions L2 and L3, it can further be shown that there is a unique periodic orbit, which attracts nearby orbits. [cf. Hale, Theorem 1.6, page 60, [4]]

5.5. The Fundamental Theorem for ω -limit sets in \mathbb{R}^2

The fundamental property of an ω -limit set is that it consists of one whole orbit, or that it is the union of more than one whole orbit. The simplest situations are

- $\omega(a)$ is an equilibrium point, i.e., the system approaches an equilibrium state as $t \rightarrow +\infty$,
- $\omega(a)$ is a periodic orbit, i.e., the system approaches an oscillatory steady state as $t \rightarrow +\infty$.

In order to motivate the fundamental theorem for ω -limit sets in \mathbb{R}^2 , we begin by discussing an example in which the ω -limit set is the union of two orbits, giving an example of a *cycle-graph*.

Consider the DE

$$x'_1 = x_2, \quad x'_2 = 2x_1 - 3x_1^2 - \mu x_2(x_1^3 - x_1^2 + \frac{1}{2}x_2^2), \quad (90)$$

For $\mu = 0$, this reduces to a Hamiltonian DE with

$$H(x_1, x_2) = \frac{1}{2}x_2^2 + x_1^3 - x_1^2, \quad (91)$$

We wish to sketch the portrait of the orbits when μ is positive but close to 0, so that the DE (90) can be thought of as a perturbation of the Hamiltonian DE. The portrait for the case $\mu = 0$ is given in Fig. 14.

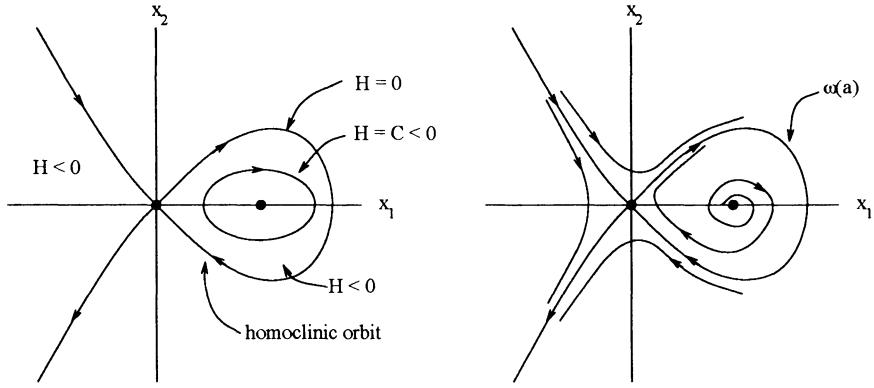


Figure 14. Phase portraits for $\mu = 0$ and $\mu \neq 0$.

We use H as given by equation (91) as a Liapunov function for the DE (90). It follows that

$$\dot{H}(x_1, x_2) = \nabla H \cdot f = -2\mu x_2^2 H(x_1, x_2). \quad (92)$$

Hence, since $\dot{H} = 0$ when $H = 0$, the level set $H = 0$ consists of orbits of the DE (as is the case when $\mu = 0$). Thus the set $S = \{(x_1, x_2) | H \leq 0, x_1 \geq 0\}$ is a closed and bounded positively invariant set. Also, it follows from equation (92) that if $\mu > 0$, then $\dot{H} \geq 0$ in S and hence the orbits in S cross the level sets $H(x_1, x_2) = C < 0$ in the outward direction. Further we can apply the Global Liapunov theorem to H on the set S to conclude that for any initial state a in the interior of S (except for the equilibrium point), the ω -limit set is the union of the homoclinic orbit and the equilibrium point $(0, 0)$. This justifies the second phase portrait in Fig. 14. This ω -limit set, which is the union of two orbits, is an example of a *cycle-graph*.

Comment: It is of interest to describe the behaviour of a solution which corresponds to the above orbit $\gamma(a)$, as $t \rightarrow +\infty$. The functions $x_1(t), x_2(t)$ are not asymptotically periodic. The reason for this is that at an equilibrium point, the velocity of the point $(x_1(t), x_2(t))$ in state space (i.e., the vector field f) is zero, and thus $(x_1(t), x_2(t))$ lingers near $(0, 0)$ for successively longer time intervals as $t \rightarrow +\infty$.

5.6. The Fundamental Theorem

Let us give some more examples of invariant sets (i.e., unions of orbits) which can arise as ω -limit sets in \mathbb{R}^2 . (See Fig. 15). and some invariant sets which *cannot* arise as ω -limit sets. (See Fig. 16). The four invariant sets in Fig. 15 are examples of cycle graphs.

Definition. A cycle graph S of a DE $x' = f(x)$ in \mathbb{R}^2 is a connected union of orbits such that

1. for all $x \in S$, $\omega(x) = \{p\}$ and $\alpha(x) = \{q\}$, where p and q are equilibrium points in S .
2. for all equilibrium points $p \in S$, there exists points $x, y \in S$ such that $\omega(x) = \{p\}$, $\alpha(y) = \{q\}$, and the number of equilibrium points in S is finite.

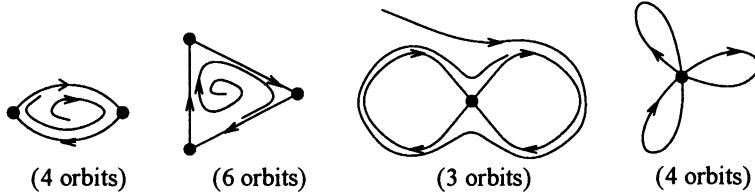


Figure 15. Examples of unions of orbits which can arise as ω -limit sets in \mathbb{R}^2 .

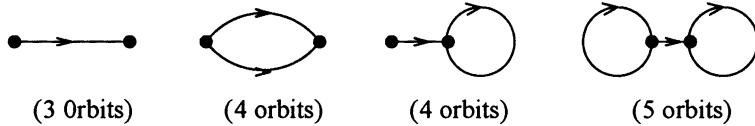


Figure 16. Examples of unions of orbits which *cannot* arise as ω -limit sets in \mathbb{R}^2 .

3. the orientations of the orbits define a continuous closed path in S .

Comment: In 1 and 2, $\alpha(x)$ denotes the α -limit set of the point x , which is the set of past limit points of x , i.e.,

$$\alpha(x) = \left\{ y \mid y = \lim_{n \rightarrow \infty} g^{t_n} x, \text{ and } t_n \xrightarrow{n \rightarrow \infty} -\infty \right\}. \quad (93)$$

We can now state:

Theorem. Consider a DE $x' = f(x)$ in \mathbb{R}^2 . Let $a \in \mathbb{R}^2$ be an initial point such that $\{g^t a \mid t \geq 0\}$ lies in a closed bounded subset $K \subset \mathbb{R}^2$. If K contains only a finite number of equilibrium points then one of the following holds:

1. $\omega(a)$ is an equilibrium point
2. $\omega(a)$ is a periodic orbit
3. $\omega(a)$ is a cycle graph.

Proof. The proof is based on the fundamental lemma of ω -limit sets in \mathbb{R}^2 . [cf. Hale, Theorem 1.3, page 230, [4]] also [cf. Lefshetz, page 230, [7]]. \square

Comment: This theorem does not generalize to DEs in \mathbb{R}^n , $n \geq 3$, or to DEs on the 2-torus. Indeed, the problem of describing all possible ω -limit sets in \mathbb{R}^n , $n \geq 3$, is presently unsolved.

6. STRUCTURAL STABILITY AND BIFURCATION THEORY

6.1. Structural Stability

In the theory of DEs, the word “stability” arises in two contexts, distinguished by the names *Liapunov Stability* and *Structural Stability*. *Liapunov Stability Question:* If a given physical system is perturbed from an equilibrium state, or from an oscillatory

steady state, does the system remain close? Mathematically, one is concerned with the behaviour of orbits in a neighborhood of an equilibrium point, or of a periodic orbit.

On the other hand, the second concept arises from the *Structural Stability Question*: Consider the DE $x' = f(x)$ in \mathbb{R}^n , with flow g^t . Suppose that the vector field $f(x)$ is perturbed, giving a vector field $\tilde{f}(x)$, and the DE $x' = \tilde{f}(x)$, with flow \tilde{g}^t . Is the flow \tilde{g}^t topologically equivalent to the flow g^t ? In particular is the long-term behaviour of solutions of the two DEs the same? If the answer is NO, we say that the given vector field is *structurally unstable* (formal definition to follow).

We now give three examples of vector fields which are structurally unstable, thereby motivating three necessary conditions for structural instability.

Example.

$$\text{Given DE: } x'_1 = x_1^2, \quad x'_2 = -x^2 \quad (\text{one singular point})$$

$$\text{Perturbed DE: } x'_1 = -\mu + x_1^2, \quad x'_2 = -x^2 \text{ with } \mu > 0 \quad (\text{two singular points})$$

As $\mu \rightarrow 0^+$ the flow \tilde{g}^t of the perturbed DE reduces to the flow g^t of the given DE. However for $\mu > 0$, \tilde{g}^t is not topologically equivalent to g^t . This shows that *the given vector field (DE) is not structurally stable*. The instability is due to the fact that *the equilibrium point $(0,0)$ of the given DE is not hyperbolic*, i.e., $Df(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$.

Example.

$$\text{Given DE: } r' = -r(r^2 - 1)^2, \quad \theta' = -1$$

$$\text{Perturbed DE: } r' = [\mu - (r^2 - 1)^2]r, \quad \theta' = -1 \text{ with } \mu > 0,$$

in polar coordinates (r, θ) . The flow \tilde{g}^t of the perturbed DE is not topologically equivalent to the flow g^t of the given DE (since \tilde{g}^t has two periodic orbits, and g^t has only one). This shows that the given vector field (DE) is not structurally stable. The instability is due to the fact that *the periodic orbit $r = 1$ does not attract or repel orbits in some neighborhoods*.

Example.

$$\text{Given DE: } x_1 = -x_1^2 + 1, \quad x'_2 = 2x_1x_2$$

$$\text{Perturbed DE: } x_1 = -x_1^2 + 1, \quad x'_2 = 2x_1x_2 - \mu(1 - x_1^2), \text{ with } \mu > 0.$$

The orbit which “joins” the two saddle points in the phase portrait of the given DE is called a *saddle connection*. Since the saddle connection is broken when $\mu > 0$, the flow of the perturbed DE is not topologically equivalent to the flow of the given DE. This shows that the given vector field (DE) is not structurally stable. The instability is due to *the presence of the saddle connection*.

6.2. Definition of Structural Stability

The statement “a vector field \tilde{f} is a perturbation of a vector field f ” means that the difference $\tilde{f} - f$ is small in some sense. In order to make this precise, one needs to define the *norm of a vector field*. For simplicity, we restrict our considerations to a bounded subset of \mathbb{R}^n .

Let $C^1(\mathcal{D})$ denote the set of C^1 vector fields which are defined on the subset $\mathcal{D} \subset \mathbb{R}^n$, where $\mathcal{D} = \{x \in \mathbb{R}^n \mid \|x\| \leq R\}$, i.e., the solid sphere of radius R . We define a norm on $C^1(\mathcal{D})$, called the *C^1 -norm*, by

$$\|f\|_1 = \max_{x \in \mathcal{D}} \|f\| + \max_{x \in \mathcal{D}} \|Df(x)\|,$$

where

$$\|f\| = \max_{1 \leq j \leq n} |f_j(x)|; \quad \|Df(x)\| = \max_{1 \leq j, k \leq n} \left| \frac{\partial f_j(x)}{\partial x_k} \right|.$$

We now use this norm to define structural stability.

We restrict our consideration to vector fields in $C^1(\mathcal{D})$ which point inwards on the boundary of \mathcal{D} , so that \mathcal{D} is a positively invariant set for the DE $x' = f(x)$. It is not essential that \mathcal{D} be a solid sphere. We could use any subset of \mathbb{R}^n that is homeomorphic to \mathcal{D} .

Definition. A vector field $f \in C^1(\mathcal{D})$ which points inwards on the boundary of \mathcal{D} is said to be structurally stable on \mathcal{D} if there exists an $\epsilon > 0$ such that for all vector fields $\tilde{f}(x) \in C^1(\mathcal{D})$, $\|f - \tilde{f}\| < \epsilon$, then the flow determined by \tilde{f} is topologically equivalent to the flow determined by f .

6.3. Structural Stability in 2-D

The previous three examples suggest that the following conditions are necessary for $f(x) \in C^1(\mathcal{D})$ to be structurally stable on \mathcal{D} :

SS1: All equilibrium points are hyperbolic

SS2: All periodic orbits are hyperbolic

SS3: There are no saddle connections.

In the above, equilibrium points etc., refer to the DE determined by the vector field, i.e., $x' = f(x)$. These three conditions are in fact also sufficient in *two dimensions* as stated in the following theorem.

Theorem. Suppose that the disc $\mathcal{D} = \{x \mid \|x\| \leq R\} \subset \mathbb{R}^2$ is a positively invariant set for the DE $x' = f(x)$. Then the vector field $f(x)$ is structurally stable on \mathcal{D} if and only if conditions SS1, SS2, and SS3 are satisfied.

Proof. [cf. Andronov and Pontrjagin, pages 247–251, [8]] □

The next question that arises is: Are “most” vector fields structurally stable, or are only a small subset structurally stable? In the 2-D case it was proved by Peixoto in 1962 that “most” vector fields are in fact structurally stable [9], where “most” has the following meaning.

Let $C_0^1(\mathcal{D})$ be the set of C^1 vector fields on $\mathcal{D} = \{x \in \mathbb{R}^2 \mid \|x\| \leq R\}$ which point inwards on the boundary of \mathcal{D} . Then we can state

Theorem. The subset $C_0^1(\mathcal{D})$ which consists of all vector fields which are structurally stable on \mathcal{D} is an open and dense subset of $C_0^1(\mathcal{D})$.

Proof. [cf. Peixoto, pages 101–120, [9]] □

Comments:

1. “Open” and “Dense” are defined in terms of the C^1 -norm in $C_0^1(\mathcal{D})$.

- We have not defined formally the term *hyperbolic periodic orbit*. A periodic orbit is hyperbolic if it attracts (or repels) all orbits in some neighborhood, at an exponential rate. This can be expressed analytically in terms of eigenvalues of a certain matrix. [cf. Hirsch and Smale, Theorem 2, page 277, [1]]. (They use the term “periodic attractor (repellor)” for hyperbolic periodic orbit.)

6.4. Bifurcations of Equilibria

Consider a DE in \mathbb{R}^n of the form $x' = f(x, \mu)$ where μ is a real parameter. Here f maps $\mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$, and hence, for each value of μ , defines a vector field on \mathbb{R}^n . Bifurcation theory, as applied to DEs, is the study of how the portrait of the orbits change as μ varies. One is interested in finding the values of μ for which a qualitative change in the orbit occurs. One thus considers values of μ for which the vector field $f(x, \mu)$ is not structurally stable.

Definition. A value μ_0 for which the vector field $f(x, \mu)$ is not structurally stable (on a suitable disc) is called a bifurcation value of μ .

The simplest bifurcations are those for which the lack of structural stability is due to the presence of a *non-hyperbolic equilibrium point*. Let us consider a simple bifurcation in one dimension that occurs when an equilibrium point has a zero eigenvalue.

Example. Consider the DE $x' = \mu x - x^3$ where $x \in \mathbb{R}$ and μ is a parameter. The equilibrium points are given by $x(\mu - x^2) = 0$ (See Fig. 17).

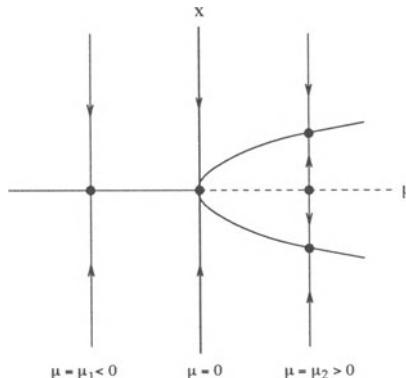


Figure 17. Bifurcation diagram for the DE $x' = \mu x - x^3$.

Comments:

- The result of this bifurcation is the creation of two new equilibrium points, and a transfer of stability to the two new points. For obvious reasons, this is called a *pitchfork bifurcation*.
- Other kinds of bifurcation in 1-D include the creation of two equilibrium points from one (called a *saddle-node bifurcation* in 2-D) [e.g., $x' = \mu - x^2$ $x \in \mathbb{R}$], or the transfer of stability between two equilibrium points (called a *transcritical bifurcation*), [e.g., $x' = \mu x - x^2$ $x \in \mathbb{R}$].

3. Often, the problem of identifying these bifurcation values in higher dimensional DEs can lead to lengthy algebra.

6.5. The Hopf Bifurcation

Consider the DE in polar coordinates $r' = (\mu - r^2)r$, $\theta' = 1$ in \mathbb{R}^2 . There is an equilibrium point at $r = 0$. For $\mu > 0$, $r = \sqrt{\mu}$ defines a periodic orbit of the DE. In addition, the linearization at the equilibrium point $r = 0$ is $Df(0, \mu) = \begin{pmatrix} \mu & 1 \\ -1 & \mu \end{pmatrix}$, which implies that $r = 0$ is an attracting spiral if $\mu < 0$ and a repelling spiral if $\mu > 0$. Also note that if $\mu \leq 0$ then $r' < 0$ for all $r > 0$. The portraits in the cases $\mu < 0$, $\mu = 0$, and $\mu > 0$ are shown in Fig. 18. The information in these portraits can be presented more concisely in a *bifurcation diagram* in the μr -plane. (See Fig. 19.) Note that $\mu = 0$ is the bifurcation value. The result of this bifurcation is 1) the creation of a stable equilibrium point, 2) the transfer of stability from an equilibrium point to a periodic orbit. This is the simplest example of a *Hopf bifurcation*.

One can rotate the previous diagram about the μ -axis (to include the angular variable θ) to obtain a bifurcation diagram which shows the actual orbits. (See Fig. 20.)

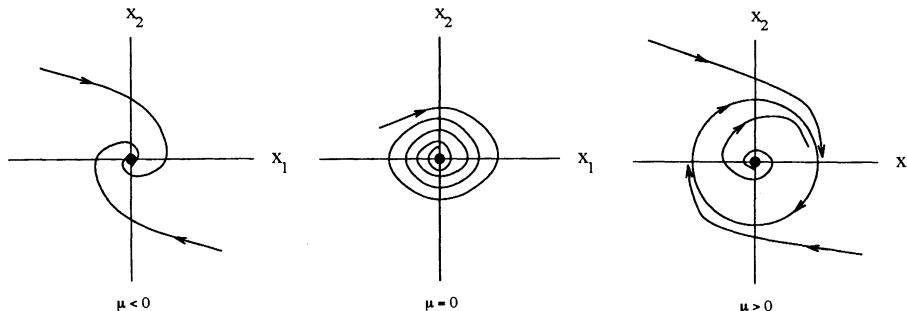


Figure 18. Phase portraits for the DE $r' = (\mu - r^2)r$ for differing values of μ .

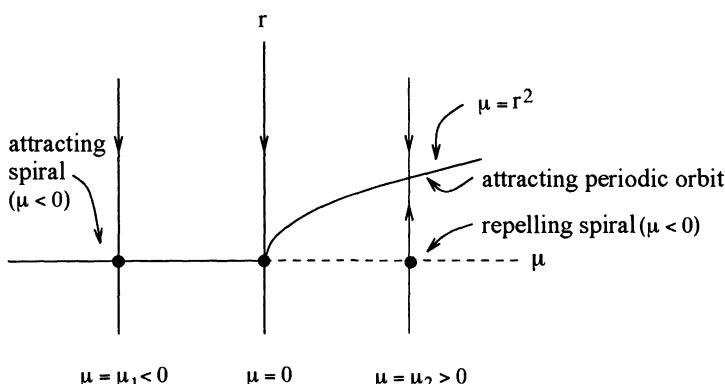


Figure 19. Bifurcation diagram for the DE $r' = (\mu - r^2)r$.

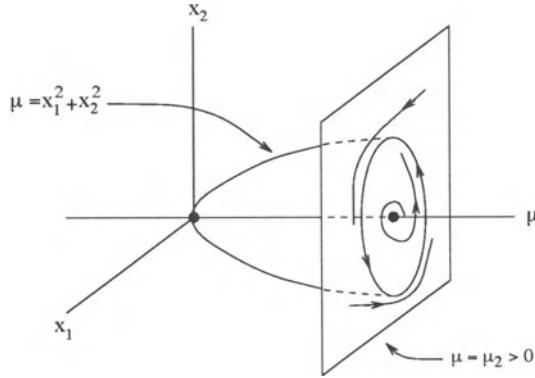


Figure 20. 3-D Bifurcation diagram for the Hopf bifurcation.

Theorem (Hopf). Consider the DE $x' = f(x, \mu)$ in \mathbb{R}^2 , where $f \in C^3$. Suppose $f(0, \mu) = 0$ for all $\mu \in I \subset \mathbb{R}$, and that $Df(0, \mu)$ has eigenvalues $\alpha(\mu) + i\beta(\mu)$. If

H1: there exists $\mu_0 \in I$ such that $\alpha(\mu_0) = 0$, $\beta(\mu_0) \neq 0$, $\alpha'(\mu_0) \neq 0$

H2: the equilibrium point $x = 0$ is not a nonlinear center when $\mu = \mu_0$

then

C: there exists a $\delta > 0$ such that for each $\mu \in (\mu_0, \mu_0 + \delta)$ or $\mu \in (\mu_0 - \delta, \mu_0)$, the DE has a unique periodic orbit (when restricted to a sufficiently small neighborhood of $x = 0$).

Proof. [cf. Hopf, vol. 94, pages 1-22 and vol. 95, pages 3-22, [10]] □

Comments:

1. The hypothesis H1 guarantees that the equilibrium point $x = 0$ is *non-hyperbolic* when $\mu = \mu_0$, and changes stability at $\mu = \mu_0$.
2. The hypothesis H2 excludes the degenerate case, in which all the periodic orbits occur at $\mu = \mu_0$, as in the linear case (for example $x'_1 = \mu x_1 + x_2$, $x'_2 = -x_1 + \mu x_2$.)

The Hopf theorem can be generalized to higher dimensions. The essential requirement is that the derivative matrix has *one pair* of pure imaginary eigenvalues and no other eigenvalues with zero real part. [cf. Guckenheimer and Holmes, page 151 [2]]

7. HIGHER DIMENSIONS

7.1. Invariant Tori and Quasiperiodic Orbits

Consider the motion of an undamped symmetric 2-mass oscillator, whose motion can be described by the DE

$$x'_1 = \omega_1 x_2, \quad x'_2 = -\omega_1 x_1, \quad x'_3 = \omega_2 x_4, \quad x'_4 = -\omega_2 x_3. \quad (94)$$

Our goal is to describe the ω -limit set $\omega(a)$ for a given initial state $a \in \mathbb{R}^4$. We let

$$x_1 = r_1 \sin \theta_1, \quad x_2 = r_1 \cos \theta_1, \quad x_3 = r_2 \sin \theta_2, \quad x_4 = r_2 \cos \theta_2, \quad (95)$$

where θ_1 and θ_2 assume values between 0 and 2π , and the values 0 and 2π are identified since they describe the same points in \mathbb{R}^4 . The DE becomes

$$r'_1 = 0, \quad \theta' = \omega_1, \quad r'_2 = 0, \quad \theta'_2 = \omega_2, \quad (96)$$

The solutions are

$$r_1 = C_1, \quad \theta = \alpha_1 + \omega_1 t, \quad r_2 = C_2, \quad \theta_2 = \alpha_2 + \omega_2 t, \quad (97)$$

where the constants $C_1, C_2, \alpha_1, \alpha_2$ are determined by the initial state. The orbits of the DE thus lie in the 2-surfaces $r_1 = C_1, r_2 = C_2$. Since these surfaces are parameterized by two variables θ_1 and θ_2 which have values modulo 2π (0 and 2π are identified), each 2-surface with $C_1 > 0$ and $C_2 > 0$ is a 2-torus. We note that these tori are invariant sets of the DE since they are unions of orbits. The nature of the orbits on each 2-torus depends on the values of the two constants ω_1 and ω_2 , which are the two natural frequencies of oscillation of the physical system. In order to illustrate this suppose that $\omega_1 = \frac{1}{4}$, $\omega_2 = 1$. (See Fig. 21.)

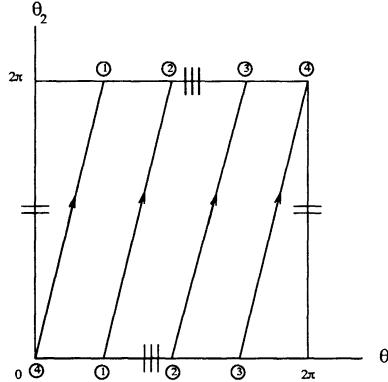


Figure 21. Plane representation of a 2-torus (= and || identified). The diagram shows an orbit corresponding to $\alpha_1 = \alpha_2 = 0$ in the solution. The points with the same number are identified, which shows that the orbit is periodic.

More generally, suppose that $\frac{\omega_1}{\omega_2} = \frac{n}{m}$, where m, n are positive integers without common factors. If $T = \frac{2\pi m}{\omega_1} = \frac{2\pi n}{\omega_2}$, then the solutions (97) satisfy

$$\theta_1(t+T) = \theta_1(t) + 2\pi m = \theta_1(t) \pmod{2\pi}$$

$$\theta_2(t+T) = \theta_2(t) + 2\pi n = \theta_2(t) \pmod{2\pi}$$

Thus the solutions are periodic of period T . The corresponding orbit on one of the invariant tori is thus periodic, and eventually closes up as it winds around the torus.

On the other hand, if $\frac{\omega_1}{\omega_2}$ is *irrational*, then the orbits are not periodic, and hence do not close up as they wind around the invariant tori. What is not immediately obvious is that as the orbit winds around the torus, it passes arbitrarily close to each point of the torus. We say that the orbit is *everywhere dense on the torus*. [cf. Arnold, [11]]

We summarize the results as follows.

Proposition 22. Consider the DE on the 2-torus defined by $\theta_1 = \omega_1$, $\theta_2 = \omega_2$.

1. if $\frac{\omega_1}{\omega_2}$ is rational, then the orbits are periodic.
2. if $\frac{\omega_1}{\omega_2}$ is irrational, then the orbits are everywhere dense on the 2-torus.

We can now draw the following conclusion concerning the ω -limit sets of the original DE (94). Consider an initial state $a = (a_1, a_2, a_3, a_4) \in \mathbb{R}^4$ and let $C_1 = a_1^2 + a_2^2$, $C_2 = a_3^2 + a_4^2$.

1. If $\frac{\omega_1}{\omega_2}$ is rational, and $C_1, C_2 > 0$, then $\omega(a)$ is the periodic orbit $\gamma(a)$ which lies on the 2-torus $r_1 = C_1$ $r_2 = C_2$.
2. If $\frac{\omega_1}{\omega_2}$ is irrational, and $C_1, C_2 > 0$, then $\omega(a)$ is the 2-torus $r_1 = C_1$ $r_2 = C_2$.

Note that in case 2, $\omega(a)$ is the union of an uncountable infinity of whole orbits, including the orbit through a .

7.2. Quasiperiodic orbits

We now discuss the type of functions which describe the solutions of a DE for which the ω -limit set is a 2-torus.

Definition. Suppose that $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}^n$ is periodic of period 2π in each argument. Suppose that $\omega_1, \omega_2 \in \mathbb{R}$ are rationally independent, i.e., $n_1\omega_1 + n_2\omega_2 \neq 0$ for all non-zero integers n_1, n_2 . Then the function $f : \mathbb{R} \rightarrow \mathbb{R}^n$ defined by $f(t) = \Psi(\omega_1 t, \omega_2 t)$ is said to be 2-quasiperiodic.

Example. Consider the DE $x' = Ax$ in \mathbb{R}^4 where

$$A = \begin{pmatrix} 0 & \omega_1 & 0 & 0 \\ -\omega_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_2 \\ 0 & 0 & -\omega_2 & 0 \end{pmatrix}.$$

Note the DE is linear and we can write the solution $x(t) = e^{tA}$ where

$$\begin{aligned} e^{tA} &= \begin{pmatrix} B(\omega_1 t) & 0 \\ 0 & B(\omega_2 t) \end{pmatrix}. \\ B(\theta) &= \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \end{aligned}$$

Thus we can write $x(t) = \Psi(\omega_1 t, \omega_2 t)a$ with $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}^4$

$$\Psi(u_1, u_2) = \begin{pmatrix} B(u_1) & 0 \\ 0 & B(u_2) \end{pmatrix}$$

and is periodic of period 2π in each argument. Thus if ω_1, ω_2 are rationally independent, and the initial state a satisfies $a_1^2 + a_2^2 \neq 0$, $a_3^2 + a_4^2 \neq 0$ then the solution $x(t)$ is a 2-quasiperiodic function. The orbit through a , $\gamma(a)$ is dense on the two torus defined by $x_1^2 + x_2^2 = a_1^2 + a_2^2$, $x_3^2 + x_4^2 = a_3^2 + a_4^2$, and is called a 2-quasiperiodic orbit.

7.3. Attracting Sets and Long-Term Behaviour

Intuitively, an attracting set is a generalization of an asymptotically stable equilibrium point or periodic orbit.

Definition. Given a DE $x' = f(x)$ in \mathbb{R}^n , a closed invariant set $A \subset \mathbb{R}^n$ is said to be an attracting set if there exists a neighborhood U of A such that

1. $g^t U \subseteq U$ for all $t \geq 0$

2. $\omega(a) \subseteq A$ for all $a \in U$,

where g^t is the flow of the DE, and $\omega(a)$ is the ω -limit set of the point a .

Definition. The basin of attraction of an attracting set A is the subset of \mathbb{R}^n defined by $\rho(A) = \{x \in \mathbb{R}^n | \omega(x) \subseteq A\}$.

If a DE has an attracting set A , then for all initial states a in the basin of attraction $\rho(A)$ the physical system approaches a “steady state” of some sort. The nature of the steady state is determined by the orbits which form the attractor. Some possibilities are summarized in the table to follow.

Attracting Sets	Long-term steady-state behaviour
Equilibrium Point	Equilibrium State
Periodic Orbit	Periodic
Invariant 2-Torus with dense orbits	2-Quasiperiodic
Invariant k-Torus with dense orbits	k-Quasiperiodic
“Strange Attractor”	“Chaotic” (none of the above)

Comment: Chaotic behaviour and Strange attractors are subjects of current research, and as yet there is no agreement on the definitions of the concepts. A Strange attractor is not a piecewise smooth surface, and can have a structure like that of a Cantor set. Chaotic behaviour occurs when neighboring orbits diverge (separate) from each other at an exponential rate, while remaining bounded, a phenomenon that is referred to as “sensitive dependence on initial conditions”. [cf. Milnor, pages 177–195 [12]; Auslander, Bhatia, and Siefert, pages 55–56 [13]]

The previous example which admits invariant 2-tori does not admit an attracting set — the invariant 2-tori do not attract neighboring orbits. Likewise, a linear DE in \mathbb{R}^2 with a centre does not admit an attracting set. But just as one can use non-linearity to create an attracting periodic orbit, one can also create an attracting 2-torus.

Example. Consider the DE in \mathbb{R}^4 ,

$$\begin{aligned} x'_1 &= \omega x_2 + x_1(\mu - r^2) \\ x'_2 &= -\omega x_1 + x_2(\mu - r^2) \\ x'_3 &= \nu x_4 + x_3(\lambda - R^2) \\ x'_4 &= \nu x_3 + x_4(\lambda - R^2) \end{aligned}$$

with $r^2 = x_1^2 + x_2^2$ and $R^2 = x_3^2 + x_4^2$, and $\omega, \nu, \mu, \lambda$ constants.

In terms of “polar coordinates”

$$x_1 = r \sin \theta, \quad x_2 = r \cos \theta, \quad x_3 = R \sin \psi, \quad x_4 = R \cos \psi,$$

the DE becomes

$$\begin{aligned} r' &= (\mu - r^2)r, & \theta' &= \omega, \\ R' &= (\lambda - R^2)R, & \psi' &= \nu. \end{aligned}$$

It follows that if $\mu > 0$ and $\lambda > 0$ then the equations $r = \sqrt{\mu}$ and $R = \sqrt{\lambda}$ define an attracting set which is an invariant 2-torus. If in addition, ω and ν are rationally independent, then the orbits on the 2-torus are dense on the 2-torus. Thus the long-term behaviour of the system would be *quasiperiodic* (with two frequencies).

The next step in understanding the long-term behaviour is to study examples which exhibit chaotic behaviour.

ACKNOWLEDGEMENTS

This manuscript was typeset in LATEX by Robert van den Hoogen.

REFERENCES

- [1] Hirsch, M.W. and Smale, S., 1974, *Differential Equations, Dynamical Systems and Linear Algebra*, (New York, Academic).
- [2] Guckenheimer, J., and Holmes, P., 1983 *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, (New York, Springer)
- [3] Nemytskii, V.V. and Stepanov, V.V., 1960, *Qualitative Theory of Differential Equations*, (Princeton, Princeton University Press).
- [4] Hale, J.K., 1969, *Ordinary Differential Equations*, (New York, Wiley).
- [5] Hartman, P., 1982, *Ordinary Differential Equations*, (New York, Wiley).
- [6] Coddington, E.A. and Levinson, N., 1955, *Theory of Ordinary Differential Equations*, (New York, McGraw-Hill).
- [7] Lefschetz, S., 1957, *Differential Equations: Geometric Theory*, (New York, Interscience).
- [8] Andronov, and Pontrjagin, 1937, *Doklady Akad. Nauk.*, **14**, 247.
- [9] Peixoto, M.M., 1962, *Topology*, **1**, 101.
- [10] Hopf, E., 1942, *Berichte Math-Phys. KI Sächs Akad. Wiss. Leipzig*, **94**, 1; 1942, *Berichte Math-Phys. KI Sächs Akad. Wiss. Leipzig*, **95**, 3.
- [11] Arnold, V.I., 1973, *Ordinary Differential Equations*, (New York, Springer).
- [12] Milnor, J., 1985, *Commun. Math. Phys.*, **99**, 177.
- [13] Auslander, J., Bhatia, N.P., and Seibert, P., 1964, *Bol. Soc. Mat. Mex.*, **9**, 55.

A SHORT COURSE ON CHAOTIC HAMILTONIAN SYSTEMS

Richard C. Churchill

Dept. of Mathematics, Hunter College and Graduate Center, CUNY
695 Park Avenue, New York, New York 10021 USA

Abstract. The lectures begin with a brief introduction to chaotic dynamics; then rapidly sketch how the ideas apply to Hamiltonian mechanics. Familiarity with Hamilton's equations, as would be encountered in elementary mechanics courses, is assumed. The treatment of geodesic flows in Section 3 is somewhat in the spirit of symplectic mechanics, but classical notation is emphasized. The presentation is rigorous only in Section 1, which is quite elementary. References for further study are provided.

1. THE NOTION OF A CHAOTIC DYNAMICAL SYSTEM

We begin with a non-Hamiltonian example, but one which easily conveys the fundamental ideas of chaotic dynamics. A few preliminaries are necessary.

Let X represent the state space of some mechanical system, and let $f : X \rightarrow X$, i.e., suppose f is a function from X into X . Define $f^0 := id_X$, where $id_X : X \rightarrow X$ denotes the identity function $x \mapsto x$, $f^1 := f$, $f^2 := f \circ f$, and, more generally, $f^n := f \circ f^{n-1}$ for all $n \geq 1$. For any $x_0 \in X$ regard $x_1 := f(x_0)$ as the state of x_0 at time 1, i.e., after one unit of time has elapsed. Then $x_n := f^n(x_0)$, for any $n \in \mathbb{N} := \{0, 1, 2, \dots\}$, is the state at time n . f^n is called the n^{th} -iterate of f , and x_n is the n^{th} -iterate of x_0 under f . The function $\varphi : \mathbb{N} \times X \rightarrow X$ determined by $(n, x) \mapsto f^n(x)$ is the *abstract dynamical system* defined by (the future or forward iterates of) f , and for any $x_0 \in X$ the collection $\mathcal{O}(x_0) := \{x_0, x_1, x_2, \dots\}$ is the (φ -)orbit of x_0 , or the orbit of x_0 under (the future or forward iterates of) f . Note that $\mathcal{O}(x_0)$ describes all future states of x_0 , and that $\dots \subset \mathcal{O}(x_n) \subset \dots \subset \mathcal{O}(x_1) \subset \mathcal{O}(x_0)$. In the subject of Dynamical Systems one is interested in the geometry of orbits, and in the relationships between nearby orbits.

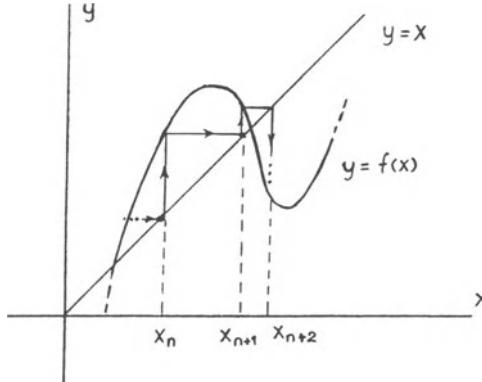


Figure 1. “Graphical analysis” of an orbit

When X is the real line \mathbb{R} , or a subset thereof, one tends to think of f in terms of a graph, and in this case it proves useful to imagine points and orbits as being on the diagonal line $y = x$ rather than on the x -axis. In particular, one replaces x with the ordered pair (x, x) , although reference is still made to x . To find x_{n+1} from x_n one moves vertically from the diagonal point x_n to the graph, and then horizontally until the diagonal is again encountered, which is easily seen to happen at the point x_{n+1} . Using this “graphical analysis” technique one can visualize (the creation of) orbits as in Figure 1.

For the example to be considered in this section we take X to be the real line \mathbb{R} , and let $\varphi : \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$ be the abstract dynamical system defined by the future iterates of $f : \mathbb{R} \rightarrow \mathbb{R}$, where

$$f(x) = \begin{cases} 3x & \text{if } x < 2/3 \\ 3x - 2 & \text{if } x \geq 2/3. \end{cases} \quad (1)$$

However, in the definitions which follow we could just as well let $f : \mathbb{R} \rightarrow \mathbb{R}$ be arbitrary, and in the first two instances replace \mathbb{R} by an arbitrary set.

1. (Fixed Points) When an orbit $\mathcal{O}(x_0)$ consists of a single point, which must necessarily be x_0 , we call that point a *fixed point* of φ and f . In the example under consideration $x_0 = 0$ and $x_0 = 1$ are the only real numbers with this property.
2. (Periodic Points and Periodic Orbits) For f as in (1) and $x_0 = 1/13$ we have $x_1 := f(x_0) = 3/13$, $x_2 := f(x_1) = f(f(x_0)) = f^2(x_0) = 9/13$, $x_3 := f(x_2) = f(f(f(x_0))) = f^3(x_0) = 1/13 = x_0$. But then $x_4 = x_1$, $x_5 = x_2$, and, in general, $x_{j+3} = x_j$ for all $j \in \mathbb{N}$. Thus $\mathcal{O}(1/13) = \{1/13, 3/13, 9/13\}$, and we say that (the orbit of) $1/13$ has *period three*. More generally, if an orbit $\mathcal{O}(x_0)$ has precisely $n \geq 2$ points and $f^n(x_0) = x_0$ then both x_0 and $\mathcal{O}(x_0)$ are said to be *periodic* (and) of period n .
3. (Points and Orbits tending to ∞) The forward iterates of a point x_0 and the orbit $\mathcal{O}(x_0)$ are said to *tend to $\pm\infty$* if $x_n \rightarrow \pm\infty$ as $n \rightarrow +\infty$; by abuse of language we also say that x_0 *tends to $\pm\infty$* . In the example under consideration take $x_0 = 1 + \varepsilon$,

where $\varepsilon > 0$ is arbitrary. Here we have $x_n := f^n(x_0) = 1 + 3^n\varepsilon \rightarrow +\infty$, and so both x_0 and $\mathcal{O}(x_0)$ tend to $+\infty$. Alternatively, take $x_0 = -\varepsilon$, where again $\varepsilon > 0$ is arbitrary. Now $x_n := f^n(x_0) = -3^n\varepsilon \rightarrow -\infty$, and so both x_0 and $\mathcal{O}(x_0)$ tend to $-\infty$.

To understand (all orbits of) φ completely all that remains is to understand orbits of points $x_0 \in (0, 1)$ (other than $x_0 = 1/13$, which is not worth setting apart in the discussion which follows). To this end it proves useful to codify (i.e., label) such x_0 as infinite sequences $\varepsilon_1\varepsilon_2\varepsilon_3\dots$ in which each ε_j is either 0, 1, or 2. In describing our method we also allow $x_0 = 0$ and 1.

Choose $x \in I := [0, 1]$ and write I as $I_0 \cup I_1 \cup I_2$, where $I_0 := [0, 1/3]$, $I_1 := (1/3, 2/3)$, and $I_2 := [2/3, 1]$. Then $x \in I_j$ for precisely one j , and we set $\varepsilon_1 := j$. (Examples: For $x = 1/13$ we have $\varepsilon_1 = 0$, since $x \in I_0$, whereas for $x = 2/3$ we have $\varepsilon_1 = 2$ since $x \in I_2$.) Having chosen ε_1 write I_{ε_1} as $I_{\varepsilon_10} \cup I_{\varepsilon_11} \cup I_{\varepsilon_12}$, where $I_{\varepsilon_1j} := [(\varepsilon_1/3) + (j/3^2), (\varepsilon_1/3) + ((j+1)/3^2)]$, $j = 0, 1, 2$ (endpoints excluded for $j = 1$). Then $x \in I_{\varepsilon_1j}$ for precisely one j , and we set $\varepsilon_2 := j$. (Examples: for $x = 1/13$ we have $\varepsilon_2 = 0$, since $x \in I_{00}$, and for $x = 2/3$ we also have $\varepsilon_2 = 0$, now because $x \in I_{20}$. Having chosen ε_1 and ε_2 write $I_{\varepsilon_1\varepsilon_2}$ as $I_{\varepsilon_1\varepsilon_20} \cup I_{\varepsilon_1\varepsilon_21} \cup I_{\varepsilon_1\varepsilon_22}$, where \dots . The association between the resulting infinite sequence $\varepsilon_1\varepsilon_2\varepsilon_3\dots$ and x will be indicated by writing

$$x \sim \varepsilon_1\varepsilon_2\varepsilon_3\dots \quad (2)$$

For example,

$$1/13 \sim 002002002002002002002002002\dots, \quad (3)$$

which the interested reader is invited to verify.

Proposition 1.1 Suppose $x \in [0, 1]$ and $x \sim \varepsilon_1\varepsilon_2\varepsilon_3\dots$. Then:

1. $x = \sum_{n=1}^{\infty} (\varepsilon_n/3^n)$.
2. $f(x) = \begin{cases} \sum_{n=1}^{\infty} (\varepsilon_{n+1}/3^n) & \text{if } \varepsilon_1 = 0 \text{ or } 2 \\ 1 + \sum_{n=1}^{\infty} (\varepsilon_{n+1}/3^n) & \text{if } \varepsilon_1 = 1 \end{cases}$.

In particular, $f(x) > 1$ if $\varepsilon_1 = 1$ and $\varepsilon_n \neq 0$ for at least one $n > 1$.

3. Suppose $y \in [0, 1]$ and $y \sim \varepsilon'_1\varepsilon'_2\varepsilon'_3\dots$. If $\varepsilon_n = \varepsilon'_n$ for $n = 1, \dots, k+1$ then $|x-y| < 3^{-k}$, whereas $|x-y| \geq 1/3$ if $\varepsilon_1 \neq \varepsilon'_1$ and $\varepsilon_1 \neq 1 \neq \varepsilon'_1$.

The initial assertion can be slightly misleading: although for a given $x \in [0, 1]$ our algorithm chooses the ε_j uniquely, one can write $x = \sum_{n=1}^{\infty} (\varepsilon_n/3^n)$, where $\varepsilon_n = 0, 1$, or 2, in more than one way.

Proof:

1. For each n we have $x \in I_{\varepsilon_1\dots\varepsilon_n}$, i.e., $(\varepsilon_n/3) + (\varepsilon_n/3^2) + \dots + (\varepsilon_n/3^n) \leq x \leq (\varepsilon_n/3^2) + \dots + (\varepsilon_n/3^n) + 1/3^n$. Now let $n \rightarrow \infty$.

2. From 1. and (1) we have

$$f(x) = \begin{cases} 3 \cdot (\sum_{n=1}^{\infty} (\varepsilon_n / 3^n)) & \text{if } \varepsilon_1 = 0 \text{ or } 1 \\ 3 \cdot (\sum_{n=1}^{\infty} (\varepsilon_n / 3^n)) - 2 & \text{if } \varepsilon_1 = 2. \end{cases}$$

For $\varepsilon_1 = 0$ this reduces to

$$3 \cdot \left(\sum_{n=2}^{\infty} (\varepsilon_n / 3^n) \right) = \sum_{n=2}^{\infty} (\varepsilon_n / 3^{n-1}) = \sum_{n=1}^{\infty} (\varepsilon_{n+1} / 3^n);$$

for $\varepsilon_1 = 1$ to

$$3 \cdot \left(1/3 + \sum_{n=2}^{\infty} (\varepsilon_n / 3^n) \right) = 1 + \sum_{n=2}^{\infty} (\varepsilon_n / 3^{n-1}) = 1 + \sum_{n=1}^{\infty} (\varepsilon_{n+1} / 3^n);$$

↗

for $\varepsilon_1 = 2$ to

$$3 \cdot \left(2/3 + \sum_{n=2}^{\infty} (\varepsilon_n / 3^n) \right) - 2 = \sum_{n=2}^{\infty} (\varepsilon_n / 3^{n-1}) = \sum_{n=1}^{\infty} (\varepsilon_{n+1} / 3^n).$$

3. In the first case the sequence construction guarantees that $x, y \in I_{\varepsilon_1 \dots \varepsilon_{k+1}}$, which has length 3^{-k} . In the second case either $x \in I_0$ and $y \in I_2$ or vice-versa. Q.E.D.

A more practical way to express the second assertion of Proposition 1.1 is:

$$x \sim \varepsilon_1 \varepsilon_2 \varepsilon_3 \dots \Rightarrow \begin{cases} f(x) \sim \varepsilon_2 \varepsilon_3 \varepsilon_4 \dots & \text{if } \varepsilon_1 \neq 1 \\ f(x) = 1 & \text{if } \varepsilon_1 = 1 \text{ and } \varepsilon_n = 0 \text{ for all } n > 1 \\ f(x) > 1 & \text{if } \varepsilon_1 = 1 \text{ and } \varepsilon_n \neq 0 \text{ for some } n > 1. \end{cases}$$

Corollary 1.2 Suppose $x_0 \sim \varepsilon_1 \varepsilon_2 \varepsilon_3 \dots \in [0, 1]$.

1. Assume for some $n \geq 1$ that $\varepsilon_n = 1$ and $\varepsilon_j = 0$ for all $j > n$. Then $x_{n+j} = f^{n+j}(x_0) = 1$ for all $j \geq 0$.
2. Assume for some $n \geq 1$ that $\varepsilon_n = 1$ but $\varepsilon_j \neq 0$ for at least one $j > n$. Then $\mathcal{O}(x_0)$ goes to ∞ .
3. Assume $\varepsilon_n \neq 1$ for all n . Then for $x_1 = f(x_0), x_2 = f(x_1) = f^2(x_0), \dots$ we have

$$\begin{aligned} x_1 &\sim \varepsilon_2 \varepsilon_3 \varepsilon_4 \dots, \\ x_2 &\sim \varepsilon_3 \varepsilon_4 \varepsilon_5 \dots, \\ x_3 &\sim \varepsilon_4 \varepsilon_5 \varepsilon_6 \dots, \\ &\vdots \end{aligned}$$

i.e., the successive iterates of x_0 are obtained from $\varepsilon_1 \varepsilon_2 \varepsilon_3 \dots$ simply by peeling off the ε_n from the left, one at a time.

The third assertion shows that when no ε_n is equal to 1 the labelling $x_0 \sim \varepsilon_1 \varepsilon_2 \varepsilon_3 \dots$ describes the itinerary of x_0 . This often makes the nature of $\mathcal{O}(x_0)$ easy to understand. For example, an independent verification of the period three property of $\mathcal{O}(1/13)$ can now be seen from (3).

Let $C := \{x \sim \varepsilon_1 \varepsilon_2 \varepsilon_3 \dots \in [0, 1] : \varepsilon_n \neq 1 \text{ for all } n\}$; this is the *Cantor set*. In view of Corollary 1.2 we have $f(C) \subset C$, and so the forward iterates of the restriction $f|C$ define an abstract dynamical system $\varphi_S : \mathbb{N} \times C \rightarrow C$, i.e., φ_S is given by $(n, x) \mapsto f^n(x)$. We will show that this “subdynamical system” of φ is the “chaotic” portion of φ .

Proposition 1.3 *The abstract dynamical system $\varphi_S : \mathbb{N} \times C \rightarrow C$ has the following “chaotic” properties.*

1. (The periodic points are dense.) *For each $x \in C$ and each $\varepsilon > 0$ there is a periodic point $y \in C$ such that $|x - y| < \varepsilon$.*
2. (There is a transitive orbit.) *There is a point $x_0 \in C$ with the following property: Given any $y \in C$ and any $\varepsilon > 0$ there is an $n \geq 1$ such that $|x_n - y| < \varepsilon$.*
3. (There is sensitive dependence on initial conditions.) *There is a $\delta > 0$ with the following property: For any $x_0 \in C$ and any $\varepsilon > 0$ there is a $y_0 \in C$ satisfying $|x_0 - y_0| < \varepsilon$ and $|x_n - y_n| > \delta$ for some $n > 0$.*

Proof:

1. Choose $k > 0$ so large that $3^{-k} < \varepsilon$ and suppose $x \sim \varepsilon_1 \varepsilon_2 \dots \varepsilon_k \varepsilon_{k+1} \varepsilon_{k+2} \dots$. Then $y \sim \varepsilon_1 \varepsilon_2 \dots \varepsilon_k \varepsilon_{k+1} \varepsilon_1 \varepsilon_2 \dots \varepsilon_k \varepsilon_{k+1} \varepsilon_1 \varepsilon_2 \dots$ has the required property.
2. The list 0, 2, 00, 02, 20, 22, 000, 002, … contains all sequences of 0’s and 2’s with n terms for any $n \geq 1$; let $x_0 \in C$ correspond to the concatenation, i.e., $x_0 \sim 0200022022000002\dots$. Given any $y \sim \varepsilon_1 \varepsilon_2 \dots \in C$ and $\varepsilon > 0$ choose $k > 0$ so large that $3^{-k} < \varepsilon$. Then $\varepsilon_1 \varepsilon_2 \dots \varepsilon_{k+1}$ must appear within the list, and as a result must occur in the string 0200022022000002… representing x_0 , say beginning at term n . But then $|x_n - y| < \varepsilon$.
3. Choose $k > 0$ such that $3^{-k} < \varepsilon$ and let $x_0 \sim \varepsilon_1 \varepsilon_2 \dots \varepsilon_{k+1} \varepsilon_{k+2} \varepsilon_{k+3} \dots$. Now let $y_0 \sim \varepsilon_1 \varepsilon_2 \dots \varepsilon_{k+1} \varepsilon'_{k+2} \varepsilon'_{k+3} \dots$, where $\varepsilon'_{k+j} := 2$ if $\varepsilon_{k+j} = 0$ and $\varepsilon'_{k+j} = 0$ if $\varepsilon_{k+j} = 2$, $j > 1$. Then $|x_0 - y_0| < \varepsilon$ and $|x_{k+1} - y_{k+1}| \geq 1/3$. Thus any $0 < \delta < 1/3$ will do. **Q.E.D.**

We use this result to illustrate “prescribed” chaos.

Proposition 1.4 *For any sequence $\varepsilon_1 \varepsilon_2 \varepsilon_3 \dots$ consisting only of 0’s and 2’s there is a point $x_0 \in C$ such that $x_n \in I_{\varepsilon_n}$ for all $n \geq 0$.*

Proof: Simply take $x_0 \sim \varepsilon_1 \varepsilon_2 \varepsilon_3 \dots$

Q.E.D.

It is the case with the example under consideration, and is felt (by some) to be a general truth (exceptions are known), that all chaos within an abstract dynamical

system is concentrated on a “fractal”. To define that term we need an intermediate concept.

A subset $S \subset \mathbb{R}^n$ is *self-similar* if

$$S = \bigcap_{j=1}^{\infty} S^{(j)}, \quad (4)$$

where

1. The $S^{(j)}$ are nested, i.e., $\mathbb{R}^n \supset S^{(1)} \supset S^{(2)} \supset S^{(3)} \dots$.
2. Each $S^{(j)}$ can be expressed as a finite union $S^{(j)} = \bigcup_{k=1}^{p_j} S_{jk}$.
3. Each S_{jk} is similar to some fixed (and j -independent) $S_0 \subset \mathbb{R}^n$. More precisely, there must be a real number $m_j \geq 1$ such that $m_j S_{jk} := \{m_j x : x \in S_{jk}\}$, for $k = 1, 2, \dots, p_j$, is congruent (in the sense of elementary Euclidean geometry) to S_0 .
4. $m_j \rightarrow \infty$ as $j \rightarrow \infty$.
5. The ratio $\log p_j / \log m_j$ is independent of j .

Examples 1.5

1. *The unit square* $S := \{x = (x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_j \leq 1, j = 1, 2\}$ is self-similar. Indeed, one can use $S^{(j)} := S$ for all j , $S_0 := S$, $p_1 = 1$ and for each $j > 1$ take $\{S_{jk}\}$ to be the collection of the 4^{j-1} congruent subsquares of S obtained by quartering each of the 4^{j-2} subsquares of $S^{(j-1)}$. Here $p_j = 4^{j-1}$, $m_j = 2^{j-1}$, and $\log p_j / \log m_j = 2$.
2. *The unit n -cube* $S := \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : 0 \leq x_j \leq 1, j = 1, \dots, n\}$ is self-similar. By arguing as in the case $n = 2$, but using n -cubes rather than squares, it is clear that one can take $S_0 = S^{(j)} = S$ for all j , and the S_{jk} as n -cubes of edge length $1/2^{j-1}$. Here $p_j = 2^{n(j-1)}$, $m_j = 2^{j-1}$, and $\log p_j / \log m_j = n$.
3. *The Cantor set* C is self-similar. Let $\{S_{jk}\} := \{I_{\epsilon_1 \epsilon_2 \dots \epsilon_j} : \epsilon_i \neq 1 \text{ for } 1 \leq i \leq j\}$ and set $S^{(j)} := \bigcup_k S_{jk}$. If $x \sim \epsilon_1 \epsilon_2 \epsilon_3 \dots \in C$ then: $\epsilon_1 \neq 1$ and so $x \in S^{(1)}$; $\epsilon_2 \neq 1$ and so $x \in S^{(2)}$; etc. We conclude that $C = \bigcap_{n=1}^{\infty} S^{(n)}$. Here $S^0 = [0, 1]$, $p_j = 2^j$, $m_j = 3^j$, and $\log p_j / \log m_j = \log 2 / \log 3 \simeq .6309$.

The ratio $\log p_j / \log m_j$ in our definition of a self-similar set S is called the *fractal dimension* of S . The fractal dimension of the unit square is therefore 2, and that of the unit n -cube is n , which is certainly in agreement with our intuitive notion of “dimension”. On the other hand, the fractal dimension of the Cantor set C is $\log 2 / \log 3 \simeq .6309$, and this is why the Cantor set qualifies as a “fractal”. That is, a self-similar subset of \mathbb{R}^n is a *fractal* if it has non integer fractal dimension.

We summarize our work thusfar with the following statement.

Theorem 1.6 *Let $\varphi : \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$ be defined by the forward iterates of the function f given in (1), and let $C \subset [0, 1]$ be the Cantor set. Then the following assertions hold.*

1. The restriction $\varphi_C : \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$ of φ to $\mathbb{R} \times C$ is chaotic (as described in Proposition 1.3).
2. The restriction of φ to $\mathbb{N} \times (\mathbb{R} \setminus C)$ exhibits no chaos (in the sense of Proposition 1.3). Specifically, there are no periodic or transitive orbits, and, although sensitivity in initial conditions does hold, each $x \in \mathbb{R} \setminus C$ is contained in an open interval I_x within $\mathbb{R} \setminus C$ such that the orbits of all $y \in I_x$ have the same asymptotic behavior as does $\mathcal{O}(x)$.

In particular, all the chaotic dynamics of φ occurs on the fractal C .

Although they will not be encountered in our work, an introduction to Chaos Theory would hardly be complete without some mention of the widely publicized “Julia” and “Mandelbrot” sets. These arise in connection with the abstract dynamical systems $\varphi_\mu : \mathbb{N} \times \mathbb{C} \rightarrow \mathbb{C}$ defined by the forward iterates of the “quadratic” maps $f_\mu : z \rightarrow z^2 - \mu$ of \mathbb{C} into \mathbb{C} . For fixed $\mu \in \mathbb{C}$ a periodic point z_0 of φ_μ is *repelling* if $|f^n)'(z_0)| > 1$, where n is the period of z_0 , and the *Julia set* J_μ of φ_μ is defined to be the closure of all such points. Computer experiments suggest that the restriction of φ_μ to $\mathbb{N} \times J_\mu$ is chaotic (in the sense of Proposition 1.3), and that J_μ is generally a fractal. (But not always, e.g., J_0 is the unit circle, which is not a fractal.) The *Mandelbrot set* is defined as $\{\mu \in \mathbb{C} : \text{the } \varphi_\mu\text{-orbit of } 0 \text{ does not tend to } \infty \text{ (on the Riemann sphere)}\}$.

2. ABSTRACT DYNAMICS

For our immediate purposes an *ordinary differential equation* on an open subset $U \subset \mathbb{R}^n$ will mean a function $f : U \rightarrow \mathbb{R}^n$; a *solution* (with *initial value* $p \in U$) will mean an open interval $(a, b) \subset \mathbb{R}$ containing 0 and a function $x : (a, b) \rightarrow U$, written $x = x(t)$, such that $\dot{x}(t) = f(x(t))$ (and $x(0) = p$) for all $t \in (a, b)$. We indicate such an equation (resp. initial-value problem) with the standard notation $\dot{x} = f(x)$ (resp. $\dot{x} = f(x)$, $x(0) = p$). The advantage of this viewpoint is geometric: f is regarded as a vector field on U , and a solution $x = x(t)$ as a curve with tangent vector $f(x(t))$ at $x(t)$.

Examples 2.1

1. One can always convert a higher order differential equation to the first order autonomous form $\dot{x} = f(x)$ by “adding variables.” For example, by writing y as x_1 and introducing (i.e., “adding”) the variables $x_2 := \dot{y}$ and $x_3 = t$ we can write the hypergeometric equation

$$t(1-t)\ddot{y} + (\gamma - (1+\alpha+\beta)t)\dot{y} - \alpha\beta y = 0$$

in the first order form

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= \left(\frac{\alpha\beta}{x_3(1-x_3)}\right)x_1 + \left(\frac{(1+\alpha+\beta)x_3 - \gamma}{x_3(1-x_3)}\right)x_2 \\ \dot{x}_3 &= 1.\end{aligned}$$

In particular, our definition of “ordinary differential equation” encompasses all the standard (ordinary) differential equations of mathematical physics.

2. When $U \subset \mathbb{R}^{2n}$, $x = (x_1, \dots, x_{2n}) = (q_1, \dots, q_n, p_1, \dots, p_n) = (q, p)$ and $H : U \rightarrow \mathbb{R}$ is a Hamiltonian the usual form

$$\dot{q}_j = \frac{\partial H}{\partial p_j}(q, p), \quad \dot{p}_j = -\frac{\partial H}{\partial q_j}(q, p), \quad j = 1, \dots, n,$$

of Hamilton's equations may be written in the $\dot{x} = f(x)$ format as

$$\dot{x} = JH_x(x);$$

here J denotes the $2n \times 2n$ matrix $\begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, with 0 and I being the respective $n \times n$ zero and identity matrices, and

$$\begin{aligned} H_x(x) &:= \left(\frac{\partial H}{\partial q_1}(q, p), \dots, \frac{\partial H}{\partial q_n}(q, p), \frac{\partial H}{\partial p_1}(q, p), \dots, \frac{\partial H}{\partial p_n}(q, p) \right)^\tau \\ &= \left(\frac{\partial H}{\partial q}(q, p), \frac{\partial H}{\partial p}(q, p) \right)^\tau = (H_q(q, p), H_p(q, p))^\tau, \end{aligned}$$

where the τ denotes “transpose”.

We need the following *fundamental existence theorem for ordinary differential equations*, which we state without proof. In the statement we assume all component functions of f are differentiable, with all partial derivatives continuous.

Theorem 2.2 Suppose $\dot{x} = f(x)$ is an ordinary differential equation on an open subset $U \subset \mathbb{R}^n$. Then there is a unique function $\varphi : \mathbb{R} \times U \rightarrow U$ with the following properties.

1. For each $p \in U$ the mapping $t \mapsto \varphi(t, p)$ is the unique solution of $\dot{x} = f(x)$, $x(0) = p$.
2. Suppose for each $t \in \mathbb{R}$ we define $\varphi^t : U \rightarrow U$ by $p \mapsto \varphi(t, p)$. Then
 - (a) $\varphi^0 = id_U$, and
 - (b) $\varphi^s \circ \varphi^t = \varphi^{s+t}$ for all $s, t \in \mathbb{R}$.
3. φ is differentiable.

(The statement of the theorem has been designed more for convenience than for correctness. As written the result implies that all solutions are *defined for all time*, i.e., for all $t \in \mathbb{R}$, but this is not generally true without further hypotheses.) $\varphi : \mathbb{R} \times U \rightarrow U$ is called the *flow* of $\dot{x} = f(x)$, and for any $t \in \mathbb{R}$ the function φ^t is the *time-t mapping* of φ . This mapping ties into the abstract dynamical systems of Section 1 through the following result.

Proposition 2.3 Suppose $\varphi_f : \mathbb{N} \times U \rightarrow U$ is the abstract dynamical system defined by the forward iterates of the time-1 mapping of the flow φ . Then $\varphi_f = \varphi|(\mathbb{N} \times U)$.

Proof: $\varphi_f(n, p) = (\varphi^1)^n(p) = \varphi^n(p)$ (by the second assertion of Theorem 2.2) $= \varphi(n, p)$ (by definition). Q.E.D.

The result suggests revising the definition of “abstract dynamical system” so as to incorporate those studied in Section 1 together with the flows of differential equations. To this end let S be any infinite subset of \mathbb{R} which contains 0 and 1 and is closed under addition (i.e., $s + t \in S$ whenever $s, t \in S$), e.g., $S = \mathbb{N}, \mathbb{R}$, or $\mathbb{Z} := \{0, \pm 1, \pm 2, \dots\}$ (S is for *semigroup*). Then an abstract dynamical system on a set X (*with semigroup* S) is a mapping $\varphi : S \times X \rightarrow X$ such that $\varphi^0 = \text{id}_X$ and $\varphi^s \circ \varphi^t = \varphi^{s+t}$ for all $s, t \in S$, where $\varphi^t : U \rightarrow U$ is the time- t mapping $x \mapsto \varphi(t, x)$. When X is a differentiable manifold and $\varphi : S \times X \rightarrow X$ is differentiable we speak of a *smooth (abstract) dynamical system*, or of a *smooth flow* when $S = \mathbb{R}$.

(“Smooth” is usually taken to mean “infinitely differentiable;” we are not being specific about differentiability assumptions.)

Examples 2.4

1. *The abstract dynamical systems of Section 1 are abstract dynamical systems in this new sense.* Indeed, by Proposition 2.3 we have $\varphi_f^n = \varphi^n = f^n$.
2. *The flow of a differential equation $\dot{x} = f(x)$ is a smooth abstract dynamical system.* By Theorem 2.2.
3. Suppose X is a set and $f : X \rightarrow X$ is a bijection. For $n \geq 0$ define f^n as before, and for $n > 0$ define $f^{-n} := (f^{-1})^n$. Then there is an associated abstract dynamical system $\varphi : \mathbb{Z} \times X \rightarrow X$, i.e., $(n, x) \mapsto f^n(x)$. φ is said to be *defined by the iterates of f*.
4. Let M be a smooth n -manifold with tangent bundle TM and projection $\pi : TM \rightarrow M$. If $U \subset M$ is a coordinate patch (i.e., a domain for generalized coordinates) then $\pi^{-1}(U) \approx U \times \mathbb{R}^n$, and any vector field $\mathcal{X} : M \rightarrow TM$ has a (generalized) coordinate description $x \in U \mapsto (x, f(x)) \in U \times \mathbb{R}^n$. Since $f : U \rightarrow \mathbb{R}^n$, this allows us to regard the ordinary differential equation $\dot{x} = f(x)$ on U as a local coordinate description of \mathcal{X} . Theorem 2.2 therefore applies, guaranteeing a smooth abstract dynamical system $\varphi : \mathbb{R} \times M \rightarrow M$ known as the *flow* of \mathcal{X} . When viewed in this context solutions of $\dot{x} = f(x)$ are called *integral curves* of \mathcal{X} .
5. Suppose $\varphi : S \times X \rightarrow X$ is an abstract dynamical system and $Y \subset X$ is φ -*invariant*, i.e., suppose $\varphi^s(Y) \subset Y$ for all $s \in S$. Then $\varphi(S \times Y)$ is an abstract dynamical system on Y , sometimes called a *subdynamical system* of φ . (Example: Let $\varphi : \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$ be as in Theorem 1.6 and let $Y = C$.) When $S = \mathbb{R}$ a subdynamical system is also called a *flow* on Y .
6. The situation of the preceding example arises when a differential equation $\dot{x} = f(x)$ on an open set $U \subset \mathbb{R}^n$ has an *integral* (or ‘first’ *integral, constant of the motion* or *conservation law*), i.e., when there is a non constant function $K : U \rightarrow \mathbb{R}$ such that $\frac{d}{dt}K(x(t)) \equiv 0$ for any solution $x = x(t)$ of the equation. In that case the *level surfaces* $\Sigma_{K,c} := \{x \in U : K(x) = c\}$, for any $c \in \mathbb{R}$, are φ -invariant sets, and the restriction $\varphi|(\mathbb{R} \times \Sigma_{K,c})$ defines a flow on each $\Sigma_{K,c}$. (Level surfaces are also called *level sets*, or *level curves* when $n = 2$.) If $\dot{x} = f(x)$ has functionally

independent integrals K_1, \dots, K_r , then φ induces a flow on the $(n-r)$ -dimensional surface $\Sigma_{K_1, c_1} \cap \dots \cap \Sigma_{K_r, c_r}$.

7. The Hamiltonian H of a Hamiltonian system $\dot{x} = JH_x(x)$ is always an integral, and the resulting flow therefore induces a flow on each level surface $\Sigma_{H,h}$. These level sets are the *energy surfaces* of H , or of $\dot{x} = JH_x(x)$, and when H is understood the notation $\Sigma_{H,h}$ will be abbreviated to Σ_h .
8. As a particular example of a Hamiltonian system consider the linear harmonic oscillator Hamiltonian $H(q, p) = \sum_{j=1}^n \frac{\omega_j}{2}(q_j^2 + p_j^2)$ with frequencies $\omega_j > 0$. Here the energy surfaces Σ_h , for $h > 0$, are (hyper)ellipsoids, and φ induces a flow on each such surface. In fact for this particular Hamiltonian one can say much more: the n functions $K_j : (q, p) \mapsto (q_j^2 + p_j^2)$ are independent integrals, the surfaces $\Sigma_{K_1, c_1} \cap \dots \cap \Sigma_{K_r, c_r}$ are n -dimensional tori which stratify these ellipsoids, and φ induces a flow on each such torus.
9. The “subdynamical system” terminology of Item 5 also arises when an abstract dynamical system $\varphi : S \times X \rightarrow X$ is restricted to $S' \times X$, where S' is any subset of S which contains 0 and 1 and the sum of any two of its elements. For example, one could restrict a flow $\varphi : \mathbb{R} \times X \rightarrow X$ to $\mathbb{Z} \times X$; this would be the *discrete (sub)dynamical system defined by the time-one map(ping) of φ* .
10. The subdynamical system notions of Items 5 and 9 are often combined. For example, suppose $Y \subset X$ is invariant under a flow $\varphi : \mathbb{R} \times X \rightarrow X$. Then the abstract dynamical system $\varphi|(\mathbb{Z} \times Y)$ would be considered a “subdynamical system” of φ .
11. Suppose X is a smooth manifold, $\varphi : \mathbb{R} \times X \rightarrow X$ is a smooth flow, and $\Gamma \subset X$ is a submanifold of codimension one (i.e., of one less dimension than X) which “slices” through orbits of φ as in Figure 2; such a Γ is a *local section* for φ . Now

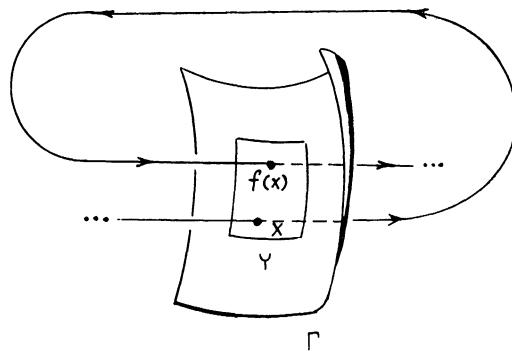


Figure 2. A local section for a flow

suppose $Y \subset \Gamma$ is such that $x \in Y$ implies $s_x \cdot x \in Y$ for some $0 < s_x \in S$, which w.l.o.g. we take to be minimal. Then the forward iterates of the mapping $f : Y \rightarrow Y$ sending $x \in Y$ to $s_x \cdot x \in Y$ define an abstract dynamical system on Y . This would also be called a “subdynamical system of φ .”

Given an abstract dynamical system $\varphi : S \times X \rightarrow X$ write $\varphi(s, x)$ as $s \cdot x$ for all $(s, x) \in S \times X$. Then the defining conditions $\varphi^0 = id_X$ and $\varphi^s \circ \varphi^t = \varphi^{s+t}$ can be written

respectively as $0 \cdot x = x$ and $s \cdot (t \cdot x) = (s + t) \cdot x$ for all $s, t \in S$ and all $x \in X$. The $(\varphi-)$ orbit of $x \in X$ is $\mathcal{O}(x) := \{s \cdot x : s \in S\}$, and the remaining definitions listed in Section 1 generalize as follows.

1. $x \in X$ is a *fixed point* (or *equilibrium point*) of φ if $\mathcal{O}(x)$ is a point.
2. $x \in X$ is a *periodic point* of φ , and $\mathcal{O}(x)$ is a *periodic orbit* of this dynamical system, if $\mathcal{O}(x) \neq \{x\}$ but $s \cdot x = x$ for some $0 < s \in S$. The smallest such $s > 0$ is the *period* of x and of $\mathcal{O}(x)$.
3. Now suppose X is a metric space (or, more generally, a topological space). An orbit $\mathcal{O}(x)$ is *transitive*, and x is called a *transitive point*, if $\mathcal{O}(x)$ is dense in X , i.e., if for each $y \in X$ and each $\varepsilon > 0$ there is an $s \in S$ such that $d(s \cdot x, y) < \varepsilon$, where $d : X \times X \rightarrow \mathbb{R}$ is the metric.
4. Again suppose X is a metric space. We say φ has sensitive dependence on initial conditions if there is a $\delta > 0$ such that for any $x \in X$ and $\varepsilon > 0$ there is a $y \in X$ and an $s \in S$ such that $d(x, y) < \varepsilon$ and $d(s \cdot x, s \cdot y) > \delta$.
5. Once more let X be a metric space. We say φ is *chaotic* if:
 - (a) the periodic points of φ are dense in X , i.e., if for each $x \in X$ and each $\varepsilon > 0$ there is a periodic point $y \in X$ satisfying $d(x, y) < \varepsilon$.
 - (b) There is a transitive orbit.
 - (c) There is sensitive dependence on initial conditions.

When one refers to a *chaotic Hamiltonian system* one means that the induced flow on some (or all) energy surface(s) is (are) chaotic.

The standard method for establishing chaos within an abstract dynamical system (Hamiltonian or not) is to realize the following example as a subsystem (in the sense of Item 11 of Examples 2.4).

Example 2.5 The Horseshoe Mapping: Let $I^2 := \{x \in \mathbb{R}^2 : 0 \leq x_j \leq 1\}$ denote the unit square of the Euclidean plane, suppose $U \subset I^2$ is an open neighborhood of I^2 , and suppose $f : U \rightarrow U$ is a diffeomorphism which carries I^2 twice “across itself” as in one of the illustrations of Figure 3.

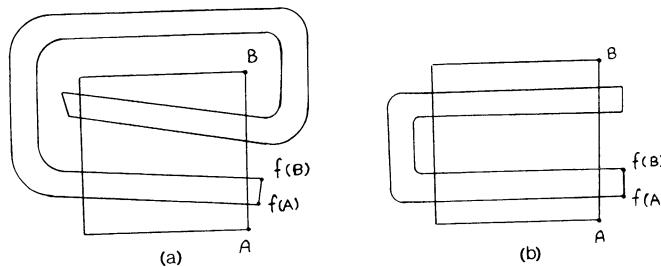


Figure 3. The Horseshoe Mapping

(Figure 3(b) explains why such an f is called a *horseshoe map(ping)*, but Figure 3(a) is more often the one which actually arises in the study of ordinary differential equations.) We will show that the abstract dynamical system $\varphi : \mathbb{Z} \times U \rightarrow U$ defined by the iterates of f has a chaotic subsystem.

First “idealize” Figure 3 as in Figure 4(a), and in that figure label the horizontal strips H_0 and H_1 as shown. The mapping f is assumed such that these strips have vertical preimages, which we label as V_0 and V_1 respectively in Figure 4(b). Now

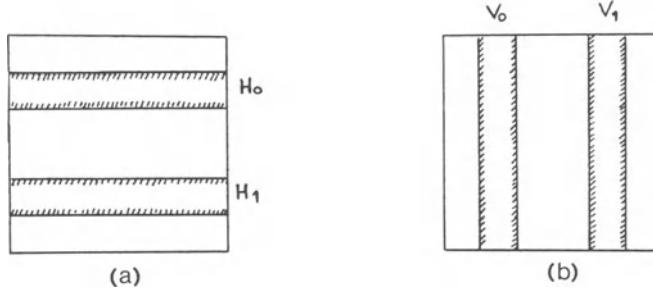


Figure 4. The idealized Horseshoe Map

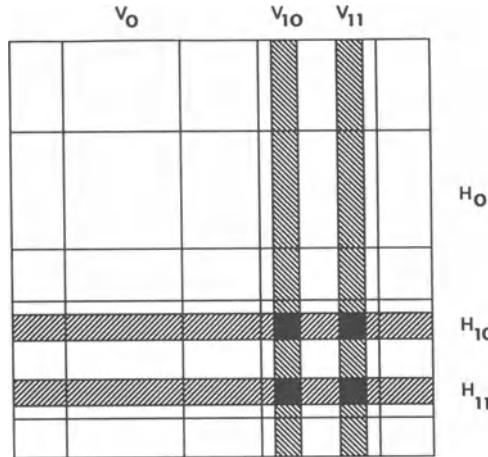


Figure 5. The nested vertical and horizontal strips

observe that $V_i \cap H_j$ will have a vertical preimage $V_{ij} \subset V_j$ and a horizontal image $H_{ij} \subset H_i$, and by then considering $V_{ij} \cap H_k$, etc., we can form a hierarchy of vertical and horizontal strips within I^2 as indicated in Figure 5. Moreover, since $\dots V_{kij} \subset V_{ij} \subset V_j$ and $\dots H_{ijk} \subset H_{ij} \subset H_i$, these strips are in nested families, so if we assume the strips “shrink” sufficiently in width, the families will intersect in vertical and horizontal line segments bisecting I^2 . Therefore, with each bi-infinite sequence

$$\dots \varepsilon_{-4} \varepsilon_{-3} \varepsilon_{-2} \varepsilon_{-1} \cdot \varepsilon_0 \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_4 \dots$$

we can associate a vertical line segment $V_{\epsilon_{-1}\epsilon_{-2}\epsilon_{-3}\epsilon_{-4}\dots}$ as well as a horizontal line segment $H_{\epsilon_0\epsilon_1\epsilon_2\epsilon_3\dots}$; hence the point x of intersection of these two segments. We write

$$x \sim \dots \epsilon_{-4}\epsilon_{-3}\epsilon_{-2}\epsilon_{-1}.\epsilon_0\epsilon_1\epsilon_2\epsilon_3\epsilon_4 \dots$$

and let $X \subset I^2$ denote the collection of all such x . Note that X is compact, being the intersection of compact sets.

We claim that X is φ -invariant and that the subsystem $\varphi|(\mathbb{Z} \times X)$ is chaotic. Indeed, as an immediate consequence of our labelling scheme we note that if $x \in X$ and $x \sim \dots \epsilon_{-4}\epsilon_{-3}\epsilon_{-2}\epsilon_{-1}.\epsilon_0\epsilon_1\epsilon_2\epsilon_3\epsilon_4 \dots$, then $f(x)$ is also in X and $f(x) \sim \dots \epsilon_{-4}\epsilon_{-3}\epsilon_{-2}.\epsilon_{-1}\epsilon_0\epsilon_1\epsilon_2\epsilon_3\epsilon_4 \dots$. Put another way: $f(x)$ is obtained from x simply by “shifting” the decimal one place to the left. (One speaks of the “Bernoulli shift” on the space of bi-infinite sequences, and of “symbolic dynamics.”) Given these remarks the claim can now be established by mimicking the proof of Theorem 1.3.

We will return to the horseshoe mapping in Section 4.

In applications sensitivity in initial conditions can be of greater concern than the density of periodic orbits. For ordinary differential equations this sensitivity is reflected in the associated *variational equations*, which we now describe.

Suppose

$$\dot{x} = f(x) \tag{5}$$

is a differential equation on an open set $U \subset \mathbb{R}^n$, $p, q \in U$, and $x_r = x_r(t)$ is the unique solution of $\dot{x} = f(x)$, $x_r(0) = r$, $r = p, q$. Then $y = y(t) := x_q(t) - x_p(t)$ measures the difference between the solutions, and Taylor’s theorem gives

$$\begin{aligned} \dot{y}(t) &= \dot{x}_q(t) - \dot{x}_p(t) = f(x_q(t)) - f(x_p(t)) \\ &= f(x_p(t) + y(t)) - f(x_p(t)) \\ &= f(x_p(t)) + \frac{df}{dx}(x_p(t))y(t) + \dots - f(x_p(t)) \\ &= \frac{df}{dx}(x_p(t))y(t) + \dots, \end{aligned}$$

where $\frac{df}{dx}(x)$ denotes the $n \times n$ Jacobian matrix of f at x . In other words, “to first order” we have

$$\dot{y}(t) = \frac{df}{dx}(x_p(t))y(t). \tag{6}$$

Equation (6) is the *variational* or *linearized equation* of $\dot{x} = f(x)$ along the solution $x = x_p(t)$, and immediately implies

$$\frac{d}{dt}|y(t)|^2 = 2\langle \dot{y}(t), y(t) \rangle = 2\langle \frac{df}{dx}(x_p(t))y(t), y(t) \rangle. \tag{7}$$

It follows, for example, that when $\frac{df}{dx}(x)$ is symmetric and positive definite the two solutions $x = x_p(t)$ and $x = x_q(t)$ must “move apart,” at least to first order, as would be expected if the system were sensitive in the initial conditions. The exponential rate

of separation (in the general case) is measured by the *Lyapunov exponent* (or *Lyapunov [order] number*)

$$\lambda := \limsup_{t \rightarrow \infty} t^{-1} \log |y(t)|,$$

which one can find discussed in standard graduate texts on ordinary differential equations (e.g., see Hartman, 1982, p. 56, or Nemytskii and Stepanov, 1960, pp. 165-183). From the definition it appears that different solutions $y = y(t)$ of (6) will yield different λ , but it is known that in general there are at most n possible values for a fixed $x = x_p(t)$. With a few notable exceptions their precise calculation is a realistic goal only when $x = x_p(t)$ is an equilibrium (i.e., constant) or periodic solution (even in the latter case there can be severe problems), and for this reason they are difficult to employ rigorously. The last assertion also applies, again with notable exceptions, to the precise calculation of (*topological*) *entropy*, which is another measure of orbit complexity. On the other hand, machine calculation of both entities can be quite helpful when attempting to understand a particular equation.

The system

$$\dot{x} = f(x), \quad \dot{y} = \frac{df}{dx}(x)y \quad (8)$$

on $U \times \mathbb{R}^n$ combines (5) and (6); a solution $(x, y) = (x(t), y(t))$ gives a solution $y = y(t)$ of the variational equation of (5) along $x = x(t)$. When φ is used to denote the flow of (5) the notation φ^T is used to denote that of (8); this is the *tangent flow* of φ . The name is explained by the following observation: If one views (5) as a local coordinate description of a vector field on a smooth n -manifold M , then (8) can be viewed as a local coordinate description of the vector field governing the induced flow on the tangent bundle TM .

In our definition of an abstract dynamical system we required $S \subset \mathbb{R}$, but this was purely for convenience. All one really needs to formulate the concept is a set X , a monoid (e.g., a group) S , and a mapping $\varphi : S \times X \rightarrow X$, which we write as $(s, x) \mapsto s \cdot x$, having the two properties $0 \cdot x = x$ and $s \cdot (t \cdot x) = (s + t) \cdot x$ for all $x \in X$ and all $s, t \in S$. Here 0 denotes the identity element of S , i.e., the unique element satisfying $0 + s = s$ for all $s \in S$.

When S is a group G the abstract systems of the previous paragraph are usually called *group actions*, or *G -actions*, and G is said to *act on X* . In this case multiplicative notation is preferred, i.e., the identity element of G is written 1 and the two defining properties become $1 \cdot x = x$ and $g \cdot (h \cdot x) = (gh) \cdot x$. The definition of the *orbit* of a point $x \in X$ is just as before, i.e., $\mathcal{O}(x) := \{g \cdot x : g \in G\}$, but the notion of a “periodic orbit” does not make sense in this generality. As an example of a group action take $G = SO(n)$, $X = \mathbb{R}^n$, and let the action be *evaluation*, i.e., suppose $g \cdot x$ is the result of applying the mapping g to the vector x . Then the orbit of any non zero $x \in \mathbb{R}^n$ is the $(n - 1)$ -sphere of radius $|x|$ centered at the origin.

When $\varphi : G \times X \rightarrow X$ is a group action on a set X the collection of G -orbits is called the *orbit space* of the action and is denoted X/G . The mapping $\pi_G : X \rightarrow X/G : x \mapsto \mathcal{O}(x)$ is the *canonical projection*; when G and X are topological spaces and φ is continuous we give X/G the quotient topology. A subset $Y \subset X$ is a *fundamental domain* for φ if each orbit intersects Y in at most two points and $\pi_G|Y : Y \rightarrow X/G$ is a

surjection. (This is weaker than the standard definition, but suffices for our purposes.) When Y has this property one can identify X/G with the quotient space Y/\sim , where for $y_1, y_2 \in Y$ we have $y_1 \sim y_2$ iff y_1 and y_2 lie on the same G -orbit. This often provides a simple way to visualize a particular orbit space.

Examples 2.6

1. Let $X = \mathbb{R}$, let G be the group generated by the translation $x \mapsto x + 1$, and let G act on \mathbb{R} by evaluation. Here $Y = [0, 1]$ is a fundamental domain, $0 \sim 1$, no other points are equivalent, and \mathbb{R}/G may therefore be regarded as the (topological) circle which results by gluing together the left and right endpoints of $[0, 1]$. We conclude that $\mathbb{R}/G \approx S^1$, where S^1 denotes the unit circle $\{z \in \mathbb{C} : |z| = 1\}$ of the complex plane.
2. Let $G = SO(n)$ act on \mathbb{R}^n by evaluation and let $Y := \{q = (q_1, 0, \dots, 0) \in \mathbb{R}^n : q_1 \geq 0\}$. Then Y intersects each G -orbit in exactly one point, and is therefore a fundamental domain. Thus $\mathbb{R}^n/G \approx Y \approx [0, \infty)$.
3. Assume $n \geq 2$ and let G be the group of Euclidean motions of \mathbb{R}^n generated by the n translations

$$(q_1, \dots, q_{j-1}, q_j, q_{j+1}, \dots, q_n) \mapsto (q_1, \dots, q_{j-1}, q_j + 1, q_{j+1}, \dots, q_n),$$

$j = 1, \dots, n$, and let G act on \mathbb{R}^n by evaluation. Then the unit cube $I^n := \{q = (q_1, \dots, q_n) \in \mathbb{R}^n : 0 \leq q_j \leq 1, j = 1, \dots, n\}$ is a fundamental domain: each orbit intersects this cube in exactly one interior point, or else in exactly two boundary points differing by a translation from G . This justifies our thinking of \mathbb{R}^n/G as the space obtained by gluing together the opposite faces of I^n . But that is a standard construction of the n -torus $T^n \approx S^1 \times \dots \times S^1$, and we conclude that $\mathbb{R}^n/G \approx T^n$.

For later purposes we note that the images of I^n under the action of G will “tile” (or *tessellate*) \mathbb{R}^n , i.e., two such images will either coincide, intersect precisely in a common face, or be disjoint, and the union of all such images will be \mathbb{R}^n .

4. Let $D := \{z \in \mathbb{C} : |z| < 1\}$ denote the open unit disc of the complex plane and let $G_D := \{g = \begin{pmatrix} \bar{p} & \bar{q} \\ q & p \end{pmatrix} : p, q \in \mathbb{C}, |p|^2 - |q|^2 = 1\}$. Then G_D acts on D , the closure of D , and on the boundary circle S^1 of D by $g \cdot z := \frac{\bar{p}z + \bar{q}}{qz + p}$. Choose any integer $k > 1$ and let $P = P_k$ denote the regular $4k$ -sided curvilinear polygon of Figure 6, in which the (Euclidean) measure of each interior angle is $2\pi/k$. It can be shown (we will not) that for each $j = 1, \dots, k$ there are elements $g_j, g'_j \in G_D$ such that g_j slides A_j onto A'_j and g'_j slides B_j onto B'_j . If $G = G_k$ is the subgroup of G_D generated by $\{g_j, g'_j\}_{j=1}^k$, then a (rather deep) theorem of Poincaré guarantees that P is a fundamental domain for the action of G , and that the images of P under this action will “tile” D in analogy with the tiling of \mathbb{R}^n described in the previous example (see, e.g., Maskit, 1971). By gluing A_j and A'_j to B_j and B'_j respectively one sees that the orbit space appears as in Figure 7; it is known mathematically as a compact (Riemann) surface of genus k . The construction generalizes to n dimensions (Maskit, 1980, pp. 73- 78).

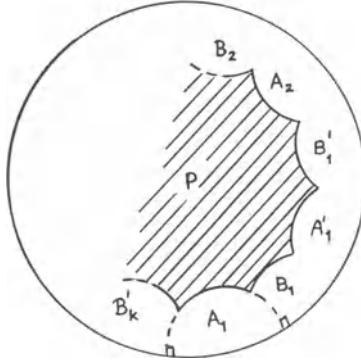


Figure 6. A regular hyperbolic polygon of $4k$ sides



Figure 7. A compact Riemann surface of genus k

3. GEODESIC FLOWS

Let $M \subset \mathbb{R}^n$ be open and let $q \mapsto S(q)$ be a Riemannian metric on M , i.e., a differentiable mapping from M into the space of positive definite symmetric $(n \times n)$ -matrices. Regard the Cartesian product $U := M \times \mathbb{R}^n$ as an open subset of $\mathbb{R}^{2n} \simeq \mathbb{R}^n \times \mathbb{R}^n \simeq \{(q, p)\}$, and define a Hamiltonian $H : U \rightarrow \mathbb{R}$ by

$$H(q, p) = \frac{1}{2} \langle (S(q))^{-1} p, p \rangle, \quad (9)$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product. If $x = x(t) = (q, p) = (q(t), p(t))$ is a solution of the corresponding Hamiltonian system

$$\dot{x} = J H_x(x), \quad (10)$$

i.e., of

$$\dot{q} = H_p(q, p), \quad \dot{p} = -H_q(q, p), \quad (11)$$

then (the image in M of) the mapping $t \mapsto q(t)$ is called a *geodesic* (in M) (*w.r.t. the Riemannian metric $S(q)$*). It occasionally proves convenient to abuse language and apply the “geodesic” terminology to (the image in U of) the mapping $t \mapsto (q(t), p(t))$ as well. When $H(q(t), p(t)) = 1/2$, i.e., when $(q(t), p(t)) \in \Sigma_{\frac{1}{2}} := \{(q, p) \in U : H(q, p) = 1/2\}$, one speaks of $t \mapsto q(t)$ (or of $t \mapsto (q(t), p(t))$) as a *unit-speed geodesic*. The *geodesic flow* on M is defined as the flow $\varphi := \varphi_S|(\mathbb{R} \times \Sigma_{\frac{1}{2}})$ on $\Sigma_{\frac{1}{2}}$, where $\varphi_S : \mathbb{R} \times U \rightarrow U$ is the

flow of (10). The terminology is misleading but standard: the “geodesic flow on M ” is not a flow on M , but a flow on $\Sigma_{\frac{1}{2}}$.

Examples 3.1

1. (The Standard Euclidean Metric) Let $M = \mathbb{R}^n$ and take $S(q)$ to be the identity matrix; this is the *standard Euclidean metric* on \mathbb{R}^n . The Hamiltonian is $H(q, p) = (1/2)\langle p, p \rangle = (1/2)|p|^2$, equations (11) become $\dot{q} = p, \dot{p} = 0$, and the flow is therefore $(t, x) = (t, (q, p)) \rightarrow (tp + q, p)$. Here geodesics are the images of the curves $t \mapsto tp + q$, i.e., they are straight lines. Unit speed geodesics are characterized by $|p| = 1$.
2. (The Jacobi Metric) Suppose $W \subset \mathbb{R}^n$ is open, $V : W \rightarrow \mathbb{R}$, $p \mapsto K(p)$ is a Riemannian metric on W , $U := W \times \mathbb{R}^n$, and a classical “kinetic plus potential energy” Hamiltonian $H : U \rightarrow \mathbb{R}$ is defined by

$$H(q, p) = \frac{1}{2}\langle (K(p))^{-1}p, p \rangle + V(q).$$

To study the resulting flow of $\Sigma_{H,h}$ let $M = M_h := \{x \in W : V(x) < h\}$, and define a Riemannian metric on M by

$$q \rightarrow 4(h - V(q))K(p).$$

This is the *Jacobi metric* associated with solutions of $\dot{x} = JH_x(x)$ of energy h , and the Hamiltonian governing the associated geodesic flow on $U := M \times \mathbb{R}^n$ is

$$H_{J,h}(q, p) = \frac{1}{4}(h - V(q))^{-1}\langle (K(p))^{-1}p, p \rangle.$$

One checks easily that $\Sigma_{H,h} \cap U = \Sigma_{H_{J,h}, \frac{1}{2}}$, and straightforward calculation gives $(H_{J,h})_x(x) = (\frac{\langle (K(p))^{-1}p, p \rangle}{4(h - V(q))^2})H_x(x)$ when $x \in \Sigma_{H_{J,h}, \frac{1}{2}}$. It follows easily that *unit speed geodesics in M w.r.t. the Jacobi metric are orientation preserving reparametrizations of solutions of $\dot{x} = JH_x(x)$ of energy h* . This example makes clear the importance of understanding geodesic flows.

3. (The Poincaré or Hyperbolic Metric) Let M denote the open unit ball $\{q \in \mathbb{R}^n : |q| < 1\}$ and take $S(q) := (\frac{4}{(1-|q|^2)^2})I$, where I denotes the $n \times n$ identity matrix; this is the *Poincaré or hyperbolic metric* on M . The geodesics are diameters of M and intersections with M of circles which meet the bounding (hyper)sphere of M orthogonally. (We will not supply a proof.)

We can construct more interesting examples by pushing the geodesics of Examples 1 and 3 down to the orbit spaces of the corresponding group actions of Examples 2.6.3 and 4. The basic idea is quite simple: if $t \mapsto \gamma(t)$ is a geodesic on M then $t \mapsto \pi \circ \gamma(t)$ will be a geodesic on M/G , where $\pi : M \rightarrow M/G$ denotes the canonical projection.

Examples 3.2

1. Take $n = 2$ in Example 2.6.3, so that $M/G \approx T^2$. In this case one can picture a geodesic on T^2 as a sequence of parallel line segments on the unit square I^2 which join to form a single curve when the edges are identified to form the torus. (See Figure 8. To obtain the parallel segments in (b) “slide” all squares in (a) back to I^2 .)

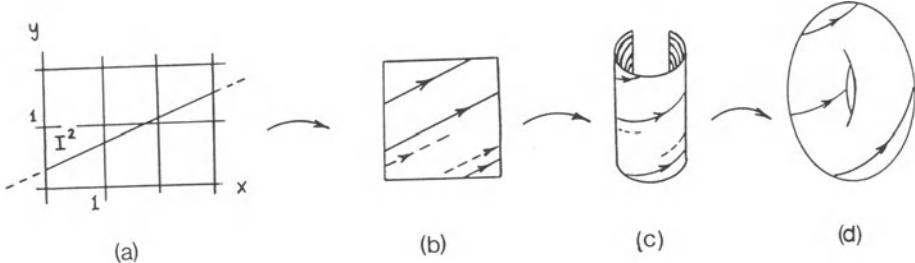


Figure 8. Geodesics on a “flat” torus

2. The analogous picture for the genus 2 ($= k$) case of Example 2.6.4 is sketched below (where considerable artistic license has been employed in replacing the 8 curvilinear edges with six straight line segments).

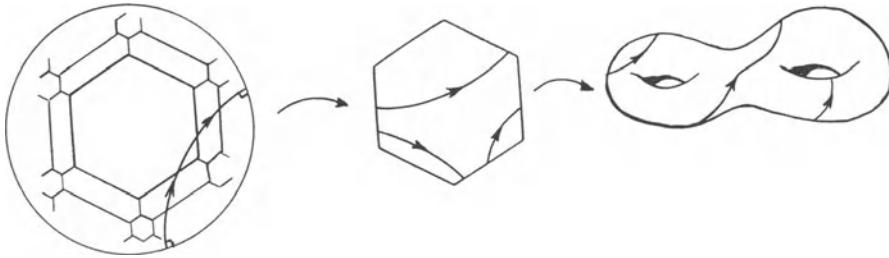


Figure 9. Geodesics on a compact surface of genus k

To describe what is meant by the “geodesic flow” on these orbit spaces we need to first reformulate the domain of H in (9), which for ease of exposition we originally assumed to be $U := M \times \mathbb{R}^n$. But to be technically correct we must take that domain to be the cotangent bundle $T^*M \approx U \times (\mathbb{R}^n)^*$ of M (because momentum vectors transform as do elements of the dual space $(\mathbb{R}^n)^*$ of \mathbb{R}^n , not as vectors in \mathbb{R}^n). From this perspective the associated geodesic equations should be regarded as a vector field on this cotangent bundle; in particular as a mapping $\mathcal{X}_H : T^*M \rightarrow T(T^*M)$, where $T(T^*M)$ denotes the tangent bundle of T^*M .

We need the fact that for the relevant group actions $\varphi : G \times M \rightarrow M$ the mappings $\varphi^g : q \mapsto g \cdot q$, for each $g \in G$, are differentiable. We denote the derivative of φ^g at $m \in M$ by $Dg(m)$ (specifically, $Dg(m) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the linear operator represented relative to the usual basis by the Jacobian matrix at m of the mapping $q \mapsto g \cdot q$) and the inverse of the adjoint (“inverse transpose”) of this operator by $(Dg(m))^{-\tau}$. We can then define the *cotangent action* $\varphi^{CT} : G \times T^*M \rightarrow T^*M$ of φ by

$$g \cdot (q, p) := (g \cdot q, (Dg(m))^{-\tau} p).$$

Finally, note that for the group actions in question we have $\dim(M) = \dim(M/G)$, and as a result the canonical projection $\pi : M \rightarrow M/G$ induces a mapping $\eta : T^*M \rightarrow T^*(M/G)$ between the relevant cotangent bundles:

$$\eta : (q, p) \mapsto (\pi(q), (Dg(m))^{-\tau} p).$$

Proposition 3.3 *Under the hypotheses of the previous three paragraphs the following assertions hold.*

1. *There is a unique function $H_R : T^*(M/G) \rightarrow \mathbb{R}$ such that $H = H_R \circ \pi$.*
2. *There is a unique vector field $\mathcal{X}_{H_R} : T^*(M/G) \rightarrow T(T^*(M/G))$ satisfying $\mathcal{X}_H = \eta^*\mathcal{X}_{H_R}$, i.e., such that $t \mapsto \pi \circ \gamma(t)$ is an integral curve of \mathcal{X}_{H_R} whenever $t \mapsto \gamma(t)$ is an integral curve of \mathcal{X}_H .*
3. *The vector field \mathcal{X}_{H_R} is Hamiltonian w.r.t. the natural symplectic structure on $T^*(M/G)$, and has H_R as Hamiltonian. In particular, H_R is an integral of the corresponding flow $\varphi_R : \mathbb{R} \times T^*(M/G) \rightarrow T^*(M/G)$.*

The subscript R in this statement is for *reduced*: the result may be viewed as a special case of *reduction*, i.e., of “reducing” a Hamiltonian system with symmetry to a smaller domain.

To prove this result one first shows that the function H in (9) is invariant under the φ^{CT} -action, which guarantees the existence of H_R ; then that the equations associated with (9) can be viewed as a local coordinate description of a vector field on $T^*(M/G)$ with the desired properties. Details are left to the reader.

By the *geodesic flow* on M/G one means the flow induced by φ_R on $\Sigma_{H_R, \frac{1}{2}} := \{x \in T^*(M/G) : H_R(x) = 1/2\}$. When $M = \mathbb{R}^n$, G is as in Example 2.6.3 and $S(q)$ is the standard Euclidean metric one speaks of the *geodesic flow on a flat torus*. When M is the open disc D of Example 2.6.4, G is the G_k of that example, and $S(q)$ is the hyperbolic metric (see Example 3.1.3) one speaks of the *geodesic flow on a compact (genus k) surface of constant negative curvature*.

Theorem 3.4 *The geodesic flow on a compact surface of constant negative curvature is chaotic.*

References to proofs are given in the Notes and Comments at the end of the lectures.

The theorem generalizes to higher dimensions, i.e., to the geodesic flow on a closed (i.e., compact and without boundary) connected manifold of negative (sectional) curvature. In fact one can prove much more than chaos for such flows: they are ergodic.

Theorem 3.4 and the higher dimensional generalizations are the fundamental examples of chaotic Hamiltonian systems; in practically all other instances where the phenomenon can be established the chaos is restricted to a subsystem. The original proof of Theorem 3.4 involved a detailed analysis of the action of the group G of Example 2.6.4 on the unit circle S^1 which forms the boundary of D ; recent proofs use different methods.

Remark: When the geodesic equations are formulated as a Lagrangian system the variational equation (6) along a solution $(q, \dot{q}) = (q(t), \dot{q}(t))$ splits into two equations; one of these, known as the *Jacobi equation*, governs variations of $q = q(t)$ alone (as opposed to variations of $(q, \dot{q}) = (q(t), \dot{q}(t))$ together). Because of a splitting of $T(T^*M)$ imposed by the metric the Jacobi equation is generally identified with a vector field on M “along $q = q(t)$ ”, known as the *Jacobi field* (see, e.g., Anosov, 1969, pp. 175-181). The “curvature” of the metric (which we have not defined, even though we have employed the terminology) appears within this equation. In particular, when the curvature is negative one finds that nearby geodesics must diverge, and as a consequence we conclude that there is a strong relationship between curvature and sensitivity in initial conditions. This suggests trying to establish sensitivity in a classical kinetic plus potential Hamiltonian system, at a fixed energy, by proving that the curvature in the Jacobi metric is negative.

4. EMBEDDINGS OF THE HORSESHOE MAPPING

As previously mentioned, embedding the Horseshoe mapping into an abstract dynamical system is the standard method for establishing chaos. The idea is most easily seen in the case of a two-degree of freedom Hamiltonian system

$$\dot{x} = JH_x(x) \quad (12)$$

defined on an open subset $U \subset \mathbb{R}^4$. (For higher dimensions, where the work is considerably more difficult; see, e.g., Wiggins, 1988, 1990.) We denote the associated flow by φ .

Suppose $\mathcal{O}(p)$ is a periodic orbit of φ of energy h (i.e., contained in $\Sigma_{H,h}$), and $\Gamma \subset \Sigma_{H,h}$ is a local section through p , i.e., a 2-dimensional surface containing p which slices through $\mathcal{O}(p)$ and all nearby orbits (see Example 2.4.11). Then there is a relatively open neighborhood $V \subset \Gamma$ of p such that for each $q \in V$ there is an $s_q > 0$ with the property that $s_q \cdot q \in \Gamma$, and we can assume s_q is minimal. The Poincaré mapping $f_\Gamma : V \rightarrow \Gamma$ of $\mathcal{O}(p)$, relative to the section Γ , is defined by $q \mapsto s_q \cdot q$ (See Figure 10).

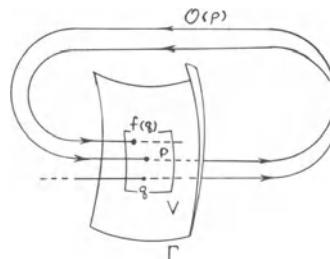


Figure 10. The Poincaré mapping of a periodic orbit

In terms of local coordinates $x = (x_1, x_2)$ on Γ vanishing at p the Poincaré mapping f_Γ assumes the form

$$f_\Gamma(x) = Ax + O(|x|^2), \quad A := \frac{df_\Gamma}{dx}(0),$$

and will be area-preserving by the Hamiltonian nature of (12). In particular $\det(A) = 1$, and as a result $\lambda \in \mathbb{C} \setminus \{0\}$ will be an eigenvalue of A iff the same holds for λ^{-1} . These eigenvalues are the *characteristic multipliers* of $\mathcal{O}(p)$, and the periodic orbit is said to be *hyperbolic* when $\lambda \in \mathbb{R} \setminus \{\pm 1\}$. (These characteristic multipliers, and as a result the definition of “hyperbolic,” depend only on the orbit $\mathcal{O}(p)$, not on p or the choice of Γ .)

STEP I: *Locate a hyperbolic periodic orbit $\mathcal{O}(p)$ of (12).*

Even this initial step may not be easy. In particular, calculating the characteristic multipliers of $\mathcal{O}(p)$ (which amounts to computing Lyapunov exponents) can prove a difficult task.

By general theory those orbits positively asymptotic to $\mathcal{O}(p)$ form a smooth manifold, known as the *stable manifold* of this orbit, and those negatively asymptotic form the *unstable manifold*. These intersect Γ in curves W^+ and W^- as indicated in Figures 11(a) and (b), and any point $q \in W^+ \cup W^-$ must be on an orbit doubly asymptotic to

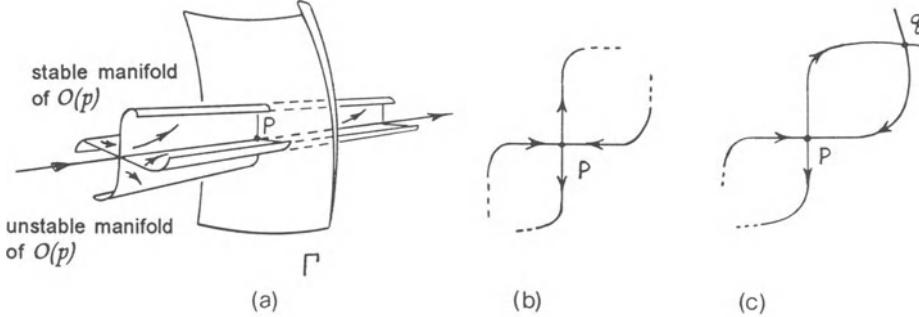


Figure 11. Stable and unstable manifolds, (a) the intersection with Γ , (b) the picture on Γ , (c) A nondegenerate homoclinic point q for p .

$\mathcal{O}(p)$; such a q is said to be a *homoclinic point* for p (w.r.t. f_Γ) or of $\mathcal{O}(p)$ (w.r.t. φ), and is said to be *nondegenerate*, or *transverse*, if the tangent vectors to W^+ and W^- at q are linearly independent (See Figure 11(c)).

STEP II: *Prove the existence of a nondegenerate homoclinic point for p .*

This can be considerably more difficult than STEP I, although for the situation under study it suffices to establish topological nondegeneracy.

Theorem 4.1 *Success with STEP I and STEP II guarantees chaos within the flow on the energy surface $\Sigma_{H,h}$ containing $\mathcal{O}(p)$.*

The idea of the proof is simple (but the details are not!). By area-preservation a small square as indicated in Figure 12 will be stretched and wrapped back upon itself under sufficiently high iterates of the mapping $f_\Gamma : V \rightarrow \Gamma$; from this one sees that the

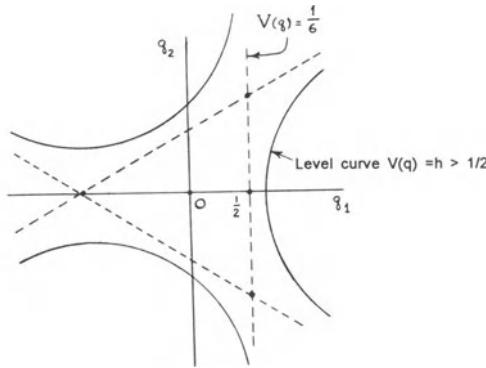


Figure 12. How the horseshoe mapping arises

Horseshoe mapping must be a subdynamical system. (In higher dimensions one must show that a small cube analogous to the square will spiral so as to intersect itself; this generally requires additional hypotheses.)

The theorem can be applied to the two-degree of freedom Hénon-Heiles Hamiltonian

$$H(q, p) = \frac{1}{2}(p_1^2 + p_2^2) + \frac{1}{2}(q_1^2 + q_2^2) + \frac{1}{3}q_1^3 - q_1 q_2^2 = \frac{1}{2}|p|^2 + V(q). \quad (13)$$

Specifically, for $h > 1/6$ the level curves of $V(q)$ appear as in Figure 13, and to satisfy

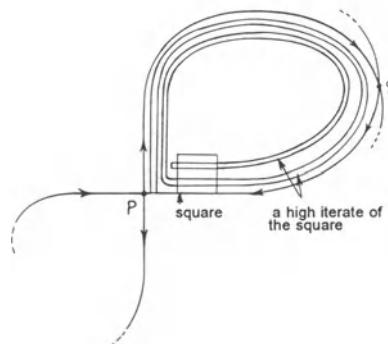


Figure 13. A typical potential curve for V , assuming $h > 1/6$

the conditions of STEP I one can use geometric methods to locate a hyperbolic orbit $\mathcal{O}(x)$ within $\Sigma_{H,h}$ which projects to a curve segment $\sigma(t)$ as indicated in Figure 14. As

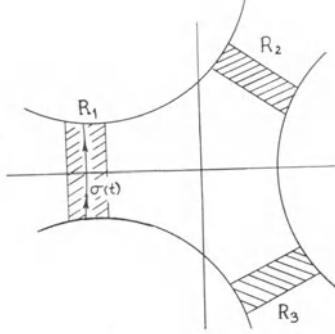


Figure 14. Projections within the potential well

a section for this orbit one uses $\Sigma_{H,h} \cap \{(q, p) : q_2 = 0 = p_2, p_1 > 0\}$, and by considerably increasing the size of h it is then possible to complete STEP II.

From the rotational symmetry of the potential one sees that there are corresponding hyperbolic orbits projecting into the other two “necks” of the potential well. The attendant horseshoe mappings can be intertwined to achieve the following “prescribed chaos” result in analogy with Proposition 1.4 (see Churchill-Rod, 1980).

Theorem 4.2 *For $h > 1/6$ sufficiently large the flow of the Hénon-Heiles Hamiltonian on $\Sigma_{H,h}$ has the following “randomness” property. Let B_j be that portion of $\Sigma_{H,h}$ projecting to the shaded region R_j of Figure 14, $j = 1, 2, 3$, and let*

$$\cdots \varepsilon_{-4} \varepsilon_{-3} \varepsilon_{-2} \varepsilon_{-1} \varepsilon_0 \varepsilon_1 \varepsilon_2 \varepsilon_3 \varepsilon_4 \cdots$$

be any bi-infinite sequence of 1's, 2's and 3's in which $\varepsilon_j \neq \varepsilon_{j+1}$ for all j . Then there is an orbit of the flow which encounters the B_j in the prescribed order. Moreover, the orbit may be assumed periodic when the sequence is periodic.

One is reminded of a magnetic pendulum, suspended above the center of an equilateral triangle, moving erratically and (therefore) unpredictably under the influence of individual magnets placed at the three vertices.

It is worth noting that when the flow for (13) on $\Sigma_{H,h}$ is formulated in terms of the Jacobi metric the associated curvature is everywhere positive (e.g., see Churchill-Pecelli-Rod, 1979, p. 99). In particular, *negative curvature in the Jacobi metric is not a necessary condition for chaos in a classical mechanical system*.

One of the standard techniques for accomplishing STEP II is the *Melnikov (perturbation) method*. One assumes that H in (12) has the form $H(q, p; \varepsilon)$, and that for $\varepsilon = 0$ there is a hyperbolic periodic orbit $\mathcal{O}(p)$ with a degenerate homoclinic point q , as in Figure 15(a). This hyperbolic periodic orbit persists for $\varepsilon > 0$ small, in the form of a hyperbolic periodic orbit $\mathcal{O}(p_\varepsilon)$, $p_\varepsilon \in \Gamma$, and the *Melnikov function* measures the (signed) distance between the associated stable and unstable manifolds, as one slides Γ along $\mathcal{O}(p_\varepsilon)$, in the spirit of Figure 15(b). STEP II is complete when this function can be shown to have a simple zero. This method was used in Holmes (1982) on a variation of (13).

NOTES AND COMMENTS

Readers wishing to pursue the ideas presented here should consult Kirchgarber and Stoffer (1990) (particularly for two degrees of freedom) or Wiggins (1988, 1990) (for higher degrees of freedom)

Section 1: There is no generally accepted definition of “chaos”; five definitions, including the one we use, are contrasted in Kirchgarber and Stoffer (1988). That given here is adapted (for my purposes) from Devaney (1989), and in Banks *et al* (1992) it is pointed out that the third condition (i.e., sensitivity) is implied by the first two.

The definition of “self-similar” is adapted (again for my purposes) from Devaney (1992); it has the advantage of permitting rapid calculation of the fractal dimension in a few simple examples. On the other hand, fractal dimension really has nothing to do with self-similarity (which seems a rather evasive concept); various approaches to “fractal dimension” are discussed in Chapter 7 of Parker and Chua (1989).

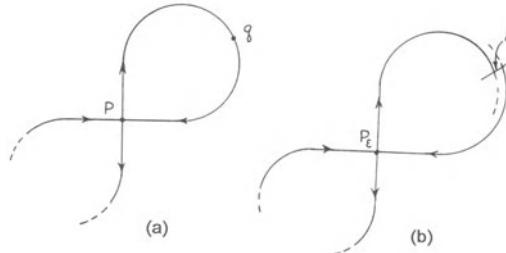


Figure 15. Melnikov’s method, (a) The $\varepsilon = 0$ orbit, (b) the $\varepsilon > 0$ orbit where the signed distance δ between stable and unstable manifolds is measured by the Melnikov function

Sections 2 and 4: The horseshoe mapping and symbolic dynamics as presented in the lectures have a long history in the subject of dynamical systems (see, e.g., pp. 628 and 629 of Birkhoff, 1968). The mapping was popularized in recent times by a paper of S. Smale (1965), and for this reason is often known as the “Smale horseshoe mapping.” The connection with chaos and non-inegrability in Hamiltonian systems was first treated in explicit general detail in Moser (1973), where one can find additional references. The work on the Hénon-Heiles Hamiltonian is from Churchill and Rod (1980) (which was written before the term “chaos” came into vogue).

For recent work on Lyapunov exponents see Oseledec (1968), Pesin (1976), and Arnold *et al* (1991).

Section 3: The original proof of Theorem 3.4 is detailed in Gottschalk and Hedlund (1955). As a detailed introduction to more recent methods see Anosov (1969); for a quick sketch see Arnol'd and Avez (1968).

ACKNOWLEDGEMENTS

I would like to thank the Department of Mathematics at the University of Calgary for providing a stimulating research environment during my Summer visit of 1993, A. Baider for suggesting that the example covered in Section 1 is well-suited for a rapid and easily comprehensible introduction to chaotic dynamics, D. L. Rod for significant help with the references, D. Hobill and R. Torrence for encouragement, I. Lagu for technical assistance, and J. Sniatycki for discussions on various concepts of chaos.

REFERENCES

- Anosov, D. V., 1969, Geodesic flows on Riemannian manifolds with negative curvature, *Proc. Steklov Math. Inst.*, vol 90 (Providence, RI, Am. Math. Soc.).
- Arnold, L. H. Crauel, J.-P. Eckmann eds., 1991, Liapunov Exponents, Proc. Oberwolfach, 1990, *Springer Lecture Notes in Mathematics*, vol 1486, (Berlin: Springer-Verlag).
- Arnol'd, V. I. and A. Avez, 1968, *Ergodic Problems of Classical Mechanics*, (New York: W. A. Benjamin).
- Banks, J., J. Brooks, G. Cairns, G. Davis and P. Stacey, 1992, On Devaney's definition of Chaos, *Am. Math. Monthly*, **99**, 332.
- Birkhoff, G. D., 1968, *Collected Mathematical Publications*, Vol. II, (New York: Dover Publications).
- Churchill, R. C., G. Pecelli and D. L. Rod, 1979, A survey of the Hénon-Heiles Hamiltonian with applications to related examples, in "Como Conference Proceedings on Stochastic Behavior in Classical and Quantum Hamiltonian Systems," *Springer Lecture Notes in Physics*, vol 99, ed G. Casati and J. Ford, (New York: Springer-Verlag).
- Churchill, R. C., and D. L. Rod, 1980, Pathology in dynamical systems III: Analytic Hamiltonians, *J. Diff. Eqns.*, **37**. 23.
- Devaney, R. L., 1989, *An introduction to Chaotic Dynamical Systems*, 2nd ed., (Redwood City, CA.: Addison-Wesley).
- Devaney, R. L. 1992, *A First Course in Chaotic Dynamical Systems, Theory and Experiment*, (Reading, MA.: Addison-Wesley).
- Gottschalk, W. H., and G. Hedlund, 1955, *Am. Math. Soc. Colloq. Publ. XXXVI*, (Providence, RI., Am. Math. Soc.).
- Hartman, P., 1982, *Ordinary Differential Equations*, 2nd ed., (Boston: Birkhäuser).
- Holmes, P., 1982, Proof of non-integrability for the Hénon-Heiles Hamiltonian near an exceptional integrable case, *Physica* **5D**, 335.
- Kirchgarber, O., and D. Stoffer, 1989, On the definition of chaos, *Z. Angew. Math., Mech.*, **69**, (1989), 175.
- Kirchgarber, O., and D. Stoffer, 1990, Chaotic behavior in simple dynamical systems, *SIAM Review*, **32**, 424.
- Maskit, B., 1971, On Poincaré's theorem for fundamental polygons, *Adv. in Math.* **7**, 219.
- Maskit, B., (1980), *Kleinian Groups*, (Berlin: Springer- Verlag).
- Moser, J., 1973, Stable and random motions in dynamical systems, *Ann. of Math. Studies*, **77**, (Princeton: Princeton University Press).

- Nemytskii, V. V., and V. V. Stepanov, 1960, *Qualitative Theory of Ordinary Differential Equations*, (Princeton: Princeton University Press).
- Oseledec, V. I., 1968, A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems, *Trans. Moscow Math. Soc.*, **19**, 197.
- Parker, T. S., and L. O. Chua, 1989, *Practical Numerical Algorithms for Chaotic Systems*, (New York: Springer-Verlag).
- Pesin, Ja. B., 1976, Characteristic Lyapunov exponents and smooth ergodic theory, *Russ. Math. Surveys*, **32**, 55.
- Smale, S., 1965, Diffeomorphisms with many periodic points, in *Differential and Combinatorial Topology*, ed S. S. Cairns (Princeton: Princeton Univ. Press), 63.
- Wiggins, S., 1988, *Global Bifurcations and Chaos, Analytical Methods*, Applied Mathematical Sciences **73**, (New York: Springer-Verlag).
- Wiggins, S., 1990, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Texts in Applied Mathematics **2**, (New York: Springer-Verlag).

GEOMETRY OF PERTURBATION THEORY

Richard Cushman

Mathematics Institute, University of Utrecht,
3508TA Utrecht, The Netherlands

Abstract. This paper gives a short exposition of the use of normal form theory and reduction to study Hamiltonian systems which are perturbations of the harmonic oscillator.¹

1. THE MATHEMATICAL PENDULUM

We begin by looking at the very simple but instructive example of the mathematical pendulum. Here phase space is \mathbf{R}^2 with canonical coordinates (x, y) . The Hamiltonian of the pendulum is

$$H(x, y) = \frac{1}{2}y^2 + (1 - \cos x). \quad (1)$$

Hamilton's equations, which describe the integral curves of the vectorfield $X_H(x, y) = y\frac{\partial}{\partial x} - \sin x\frac{\partial}{\partial y}$, are

$$\begin{aligned}\dot{x} &= y \\ \dot{y} &= -\sin x.\end{aligned}$$

The well known phase portrait of X_H is given in Figure 1. Since $X_H(0, 0) = (0, 0)$, the origin $O = (0, 0)$ is an equilibrium point of X_H . We want to study the behavior of the integral curves of X_H near O . A first step toward doing this is to linearize X_H at

¹For the expert, there is nothing new here. However, these ideas seem not have become well known to the physics community and there is no adequate source for them in the literature for the nonexpert. Essentially the results of this paper can be found in [1] and [2].

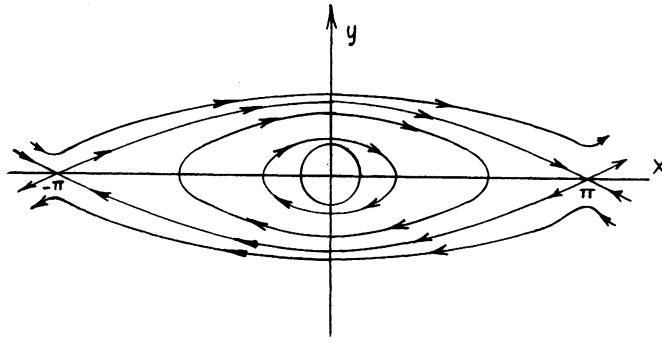


Figure 1. Phase plane portrait of the mathematical pendulum. Around the fixed point $(x, y) = (0, 0)$ are an infinite number of periodic motions. The fixed points at $(x, y) = \pm(\pi, 0)$ are unstable equilibria joined by a separatrix, outside of which the motions are free rotations.

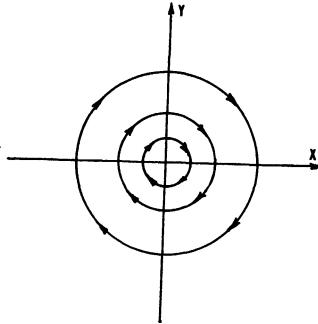


Figure 2. Flow of X_H in the vicinity of $O = (0, 0)$.

O . We obtain a new vectorfield $X_{H_2} = y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y}$, which is the Hamiltonian vectorfield of the harmonic oscillator. The flow of X_{H_2} (see Figure 2) is given by

$$\varphi_t^{H_2}(x, y) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The next question is how do we obtain a higher order approximation to the flow of X_H near O ? To answer this, we first scale the variables obtaining a new Hamiltonian

$$\begin{aligned} \widetilde{H}(x, y) &= \frac{1}{\varepsilon^2} H(x, y) = \frac{1}{2} y^2 + \frac{1}{\varepsilon^2} (1 - \cos \varepsilon x) \\ &= \frac{1}{2} (y^2 + x^2) - \frac{\varepsilon^2}{4!} x^4 + O(\varepsilon^4). \end{aligned}$$

When $\varepsilon > 0$ and small, we are looking at the Hamiltonian

$$H = H_2 + \varepsilon^2 H_4 + O(\varepsilon^4)$$

in a neighborhood of O . To avoid baroque notation we have dropped the tilde on \widetilde{H} . To find the next order approximation to the flow of X_H we average H_4 over the orbits of X_{H_2} . More formally, we must compute

$$\overline{x^4} = \frac{1}{2\pi} \int_0^{2\pi} (\varphi_t^{H_2})^* x^4 dt = \frac{1}{2\pi} \int_0^{2\pi} (x \cos t - y \sin t)^4 dt.$$

To find this integral it is better to use complex conjugate coordinates

$$z = x + iy \quad \text{and} \quad \bar{z} = x - iy$$

Then Hamilton's equations for X_{H_2} are

$$\begin{aligned}\dot{z} &= \dot{x} + i\dot{y} = y - ix = -iz = -2i\frac{\partial H_2}{\partial \bar{z}} \\ \dot{\bar{z}} &= i\bar{z} = 2i\frac{\partial H_2}{\partial z},\end{aligned}$$

where $H_2 = \frac{1}{2}z\bar{z}$. The flow of $X_{H_2} = -i(z\frac{\partial}{\partial z} - \bar{z}\frac{\partial}{\partial \bar{z}})$ is

$$\varphi_t^{H_2}(z, \bar{z}) = (z e^{-it}, \bar{z} e^{it}).$$

Therefore

$$\overline{z^4} = \frac{1}{2\pi} \int_0^{2\pi} (\operatorname{Re}(ze^{-it}))^4 dt = \frac{1}{2^4} \binom{4}{2} (z\bar{z})^2.$$

This follows after expanding the integrand using the binomial theorem, noting that the only term in the integrand which does not have an exponential factor is $(z\bar{z})^2$, and using the fact that $\int_0^{2\pi} e^{int} dt = 0$. Hence

$$\overline{H}_4 = -\frac{1}{64}(z\bar{z})^2.$$

To second order the averaged Hamiltonian is

$$\overline{H} = H_2 + \varepsilon^2 \overline{H}_4 = \frac{1}{2}z\bar{z} - \frac{\varepsilon^2}{64}(z\bar{z})^2.$$

Hamilton's equations for $X_{\overline{H}}$ are

$$\begin{aligned}\dot{z} &= -i(1 - \frac{\varepsilon^2}{16}(z\bar{z})^2)z \\ \dot{\bar{z}} &= i(1 - \frac{\varepsilon^2}{16}(z\bar{z})^2)\bar{z}.\end{aligned}$$

Since H_2 is an integral of $X_{\overline{H}}$, the flow of $X_{\overline{H}}$ in complex conjugate coordinates is

$$\varphi_t^{\overline{H}}(z, \bar{z}) = (z e^{-it(1 - \frac{\varepsilon^2}{16}(z\bar{z})^2)}, \bar{z} e^{it(1 - \frac{\varepsilon^2}{16}(z\bar{z})^2)}).$$

In geometric terms, $\varphi_t^{\overline{H}}$ is a twist map which preserves circles and rotates them through an angle varying with their radii (see Figure 3). Thus we have found a second order approximation to the flow of the mathematical pendulum near the origin.

The question arises: how does one obtain a higher order approximation? To answer this we must reexamine the averaging process. The following argument shows that to second order we can obtain the averaged Hamiltonian by a canonical change of coordinates which is the time 1 map of the flow of a Hamiltonian vectorfield. This is the basic idea of normal form theory. For the mathematical pendulum consider the Hamiltonian function

$$F(z, \bar{z}) = \frac{1}{16}(-\frac{1}{4i}z^4 - \frac{2}{i}z^3\bar{z} + \frac{2}{i}z\bar{z}^3 + \frac{1}{4i}\bar{z}^4).$$

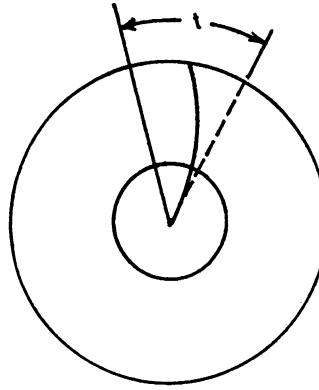


Figure 3. The twist map, φ_t^H .

To second order the time 1 map of the flow of the Hamiltonian vectorfield $X_{\epsilon^2 F}$ is

$$\varphi(z, \bar{z}) = (z, \bar{z}) + \epsilon^2 X_F(z, \bar{z}) + O(\epsilon^3).$$

Hence, after changing coordinates by φ , the Hamiltonian of the mathematical pendulum up to second order is

$$\begin{aligned} \widetilde{H}(z, \bar{z}) &= (\varphi^* H)(z, \bar{z}) = H(\varphi(z, \bar{z})) \\ &= H(z, \bar{z}) + \epsilon^2 DH(z, \bar{z})X_F(z, \bar{z}) + O(\epsilon^3), \\ &\quad \text{using the Taylor expansion of } H \\ &= H_2(z, \bar{z}) + \epsilon^2(H_4(z, \bar{z}) + DH_2(z, \bar{z})X_F(z, \bar{z})) + O(\epsilon^3), \\ &= H_2(z, \bar{z}) + \epsilon^2 \overline{H}_4(z, \bar{z}) + O(\epsilon^3), \end{aligned}$$

since

$$DH_2(z, \bar{z})X_F(z, \bar{z}) = \overline{H}_4(z, \bar{z}) - H_4(z, \bar{z}).$$

Note that after the coordinate change φ the terms of order three and higher of the new Hamiltonian \widetilde{H} are different from those of the averaged Hamiltonian of the mathematical pendulum.

Thus it makes no sense to average the entire series expansion of the Hamiltonian of the pendulum because the cubic terms and higher of this averaged Hamiltonian are destroyed by the coordinate change. Of course we can repeat the averaging process on the new Hamiltonian \widetilde{H} by applying a suitable coordinate change.

2. ALGEBRA FOR NORMAL FORMS

To do this iterated averaging efficiently, we need to be able to compute the higher order terms in the Taylor series of the flow of the Hamiltonian vectorfield and the higher order terms of the new Hamiltonian obtained after changing coordinates. The necessary algebra is encapsulated in the concept of a Poisson algebra.

Let phase space be \mathbf{R}^{2n} with canonical coordinates (x, y) , that is,

$$\{x_i, x_j\} = \{y_i, y_j\} = 0, \quad \text{and} \quad \{x_i, y_j\} = \delta_{ij},$$

where $\{ , \}$ is the usual Poisson bracket. Let \mathcal{F} be the space of formal power series Hamiltonians on \mathbf{R}^{2n} , that is, $H \in \mathcal{F}$ if and only if

$$H = H_0 + \varepsilon H_1 + \cdots + \varepsilon^n H_n + \cdots,$$

where $H_m \in \mathcal{F}_m$ and \mathcal{F}_m is the set of homogeneous polynomials of degree m on \mathbf{R}^{2n} . On \mathcal{F} define a Poisson bracket by

$$\{F, G\} = \sum_i \left(\frac{\partial F}{\partial x_i} \frac{\partial G}{\partial y_i} - \frac{\partial F}{\partial y_i} \frac{\partial G}{\partial x_i} \right). \quad (2)$$

Then $(\mathcal{F}, \{ , \})$ is a Lie algebra. On \mathcal{F} define a multiplication \bullet by

$$(F \bullet G)(x, y) = F(x, y)G(x, y).$$

Then (\mathcal{F}, \bullet) is a commutative ring with unit. In addition,

$$\{F, G \bullet H\} = \{F, G\} \bullet H + \{F, H\} \bullet G. \quad (3)$$

Thus $\mathcal{A} = (\mathcal{F}, \{ , \}, \bullet)$ is the Poisson algebra of formal Hamiltonian functions on \mathbf{R}^{2n} . All our normal form calculations will be done in \mathcal{A} .

For $F \in \mathcal{F}$ consider the linear map

$$ad_F : \mathcal{F} \longrightarrow \mathcal{F} : G \longrightarrow \{F, G\} = \sum_i \left(\frac{\partial F}{\partial x_i} \{x_i, G\} + \frac{\partial F}{\partial y_i} \{y_i, G\} \right).$$

Because of (3), ad_F is a derivation on (\mathcal{F}, \bullet) .

Example. Since

$$ad_{x_i} x_j = 0 \quad \text{and} \quad ad_{x_i} y_j = \delta_{ij},$$

it follows that ad_{x_i} is the vectorfield $\frac{\partial}{\partial y_i}$. Similarly, ad_{y_i} is the vectorfield $\frac{\partial}{\partial x_i}$. Here the coordinates x_i, y_i on \mathbf{R}^{2n} are thought of as linear polynomials.

The derivation ad_F is a formal vectorfield on \mathbf{R}^{2n} called the Hamiltonian vectorfield of F . From the example it follows that ad_F is the *negative* of the usual Hamiltonian vectorfield X_F associated to F . The formal flow of ad_F is the one parameter group of linear maps

$$\varphi^F : \mathbf{R} \longrightarrow L(\mathcal{F}, \mathcal{F}) : \varepsilon \longrightarrow \varphi_\varepsilon^F = \exp \varepsilon ad_F = \sum_{n \geq 0} \frac{\varepsilon^n}{n!} ad_F^n. \quad (4)$$

Here ad_F^n is the linear map formed by composing ad_F with itself n times. Actually, φ^F is a one parameter group of automorphisms of the Poisson algebra \mathcal{A} . In the literature, φ_ε^F is called the Lie series associated to F . For normal form theory the following property of Lie series is crucial²

$$(\varphi_\varepsilon^F)^* H = H \circ \varphi_\varepsilon^F = (\exp \varepsilon ad_F) H. \quad (5)$$

²For a proof of (4) and (5) see [3] p. 65–66.

3. NORMAL FORM

In this section we determine the normal form of a Hamiltonian which is a perturbation of the harmonic oscillator Hamiltonian.

Consider the Poisson algebra $(\tilde{\mathcal{F}}, \{ , \}, \bullet)$ of formal power series Hamiltonians on \mathbf{R}^{2n} of the form

$$H = H_2 + \varepsilon H_3 + \cdots + \varepsilon^{m-2} H_m + \cdots \quad (6)$$

where H_m is a homogeneous polynomial of degree m on \mathbf{R}^{2n} and

$$H_2(x, y) = \frac{1}{2} \sum_{i=1}^n (y_i^2 + x_i^2)$$

is the Hamiltonian of the n -degree of freedom harmonic oscillator. We say that H is in normal form³ if and only if there is a canonical change of coordinates such that in the new coordinates H is

$$\tilde{H} = H_2 + \varepsilon^2 \tilde{H}_4 + \cdots + \varepsilon^{2m} \tilde{H}_{2m} + \cdots,$$

where $\tilde{H}_{2m} \in \ker ad_{H_2}$ for every $m \geq 2$.

This is all very nice, but how do we find the canonical coordinate change which brings a given H into normal form? First we try to remove the term H_3 by a canonical coordinate change $\varphi_\varepsilon^{F_3}$ which is the time ε map of the formal flow of the formal Hamiltonian vectorfield X_{F_3} for a suitable choice of cubic homogeneous polynomial $F_3 \in \tilde{\mathcal{F}}_3$. Changing coordinates by $\varphi_\varepsilon^{F_3}$ gives a new Hamiltonian

$$\begin{aligned} H^1 &= (\varphi_\varepsilon^{F_3})^* H = (\exp \varepsilon F_3) H \\ &= H + \varepsilon ad_{F_3} H + O(\varepsilon^2) \\ &= H_2 + \varepsilon (H_3 + ad_{F_3} H_2) + O(\varepsilon^2), \text{ using (5)} \\ &= H_2 + \varepsilon (H_3 - ad_{H_2} F_3) + O(\varepsilon^2), \\ &\quad \text{since } \{F_3, H_2\} = -\{H_2, F_3\}. \end{aligned}$$

To go further we need to know something about the linear map

$$ad_{H_2} : \tilde{\mathcal{F}}_m \longrightarrow \tilde{\mathcal{F}}_m.$$

Claim. The map ad_{H_2} is semisimple, that is, every ad_{H_2} -invariant subspace of $\tilde{\mathcal{F}}_m$ has an ad_{H_2} -invariant complementary space. In particular

$$\tilde{\mathcal{F}}_m = \ker ad_{H_2} \oplus \text{im } ad_{H_2}. \quad (7)$$

Proof. It suffices to show that ad_{H_2} is diagonalizable in a suitable basis. Choose complex conjugate coordinates

$$z_j = x_j + iy_j \quad \bar{z}_j = x_j - iy_j.$$

In these coordinates H_2 becomes $\frac{1}{2} \sum_{j=1}^n z_j \bar{z}_j$ and

$$ad_{H_2} = i \sum_{j=1}^n \left(z_j \frac{\partial}{\partial z_j} - \bar{z}_j \frac{\partial}{\partial \bar{z}_j} \right).$$

³The normal form is determined up to a canonical transformation which commutes with the flow of the harmonic oscillator. For a proof of this see [4].

Using the basis of homogeneous monomials

$$z^j \bar{z}^k = z_1^{j_1} \cdots z_n^{j_n} \bar{z}_1^{k_1} \cdots \bar{z}_n^{k_n}$$

we see that

$$ad_{H_2}(z^j \bar{z}^k) = i(|j| - |k|) z^j \bar{z}^k, \quad (8)$$

where $|j| = j_1 + \cdots + j_n$. In other words, ad_{H_2} is diagonalizable.

We now find a basis for $\ker ad_{H_2}$. From (8) it follows that $z^j \bar{z}^k \in \ker ad_{H_2}$ if and only if $|j| = |k|$. Let us look at the quadratic monomials in $\ker ad_{H_2}$. Then $|j| = |k| \& |j| + |k| = 2$, that is, $|j| = |k| = 1$. Therefore the monomials $\pi_{jk} = z_j \bar{z}_k$ for $1 \leq j \leq n$ and $1 \leq k \leq n$ span $\ker ad_{H_2} \cap \tilde{\mathcal{F}}_2$. The following argument shows that $\{\pi_{jk}\}$ generate $\ker ad_{H_2}$ considered as a subalgebra of $(\tilde{\mathcal{F}}, \bullet)$. Write the factors of the monomial $z^j \bar{z}^k \in \ker ad_{H_2}$ as two lists

$$\begin{aligned} & \overbrace{z_1 z_1 \cdots z_1}^{j_1} \overbrace{z_2 \cdots z_2}^{j_2} \cdots \overbrace{z_n z_n \cdots z_n}^{j_n} \\ & \overbrace{\bar{z}_1 \cdots \bar{z}_1}^{k_1} \overbrace{\bar{z}_2 \cdots \bar{z}_2}^{k_2} \cdots \overbrace{\bar{z}_n \bar{z}_n \cdots \bar{z}_n}^{k_n}. \end{aligned}$$

Since $|j| = |k|$ these two lists have the same number of entries. Pairing off the entries expresses the monomial $z^j \bar{z}^k$ as a product of quadratic monomials in $\ker ad_{H_2}$. \square

We return to finding the normal form of H . In complex conjugate coordinates we have

$$H^1 = (\varphi_\varepsilon^{F_3})^* H = H_2 + \varepsilon(H_3 - ad_{H_2} F_3) + O(\varepsilon^2).$$

Since $H_3 \in \tilde{\mathcal{F}}_3$ and $\ker ad_{H_2} \cap \tilde{\mathcal{F}}_3 = \{0\}$, by (7) we see that $H_3 \in \text{im } ad_{H_2}$, that is, there is an $F_3 \in \tilde{\mathcal{F}}_3$ such that $ad_{H_2} F_3 = H_3$. Choosing F_3 this way we see that after the coordinate change $\varphi_\varepsilon^{F_3}$, the transformed Hamiltonian H^1 has no cubic terms. We proceed one step further in the normalization of H . Using (5) write

$$H^1 = (\exp \varepsilon ad_{F_3}) H = H_2 + \varepsilon^2 H_4^1 + O(\varepsilon^3).$$

Invoking (7), we may write $H_4^1 = \overline{H_4^1} + \widehat{H_4^1}$, where $\overline{H_4^1} \in \ker ad_{H_2} \cap \tilde{\mathcal{F}}_4$ and $\widehat{H_4^1} \in \text{im } ad_{H_2} \cap \tilde{\mathcal{F}}_4$. Choose F_4^1 so that $ad_{H_2} F_4^1 = \overline{H_4^1}$. Then

$$\begin{aligned} H^2 &= (\varphi_\varepsilon^{F_4^1})^* H^1 = (\exp \varepsilon^2 ad_{F_4^1}) H^1 \\ &= H_2 + \varepsilon^2 (H_4^1 - ad_{H_2} F_4^1) + O(\varepsilon^3) \\ &= H_2 + \varepsilon^2 (\overline{H_4^1} + \widehat{H_4^1} - ad_{H_2} F_4^1) + O(\varepsilon^3) \\ &= H_2 + \varepsilon^2 \overline{H_4^1} + O(\varepsilon^3). \end{aligned}$$

Note that the canonical coordinate change $\varphi_\varepsilon^{F_4^1}$ does not effect the terms in H^1 of order less than two. Repeating this process shows that we can find a canonical transformation which brings H into normal form.

From the definition of the normal form of H , it follows that the harmonic oscillator Hamiltonian H_2 is an integral for the normalized Hamiltonian \widetilde{H} . However, the original Hamiltonian H need not have any integral other than functions of H . Thus the normal form is more symmetric than the original Hamiltonian.

4. REDUCTION OF SYMMETRY

In this section we show how to remove the symmetry of the normalized Hamiltonian and so get a better idea of its flow. We restrict ourselves to two degrees of freedom.

Suppose that H is a smooth function on \mathbf{R}^4 which is invariant under the flow of the harmonic oscillator vectorfield X_{H_2} . Here $H_2 = \frac{1}{2} \sum_{i=1}^2 (y_i^2 + x_i^2)$. Then H is a smooth function \mathcal{H} of the Hopf variables

$$\begin{aligned} w_1 &= 2(x_1x_2 + y_1y_2) = 4 \operatorname{Re} z_1\bar{z}_1 \\ w_2 &= 2(x_1y_2 - x_2y_1) = 4 \operatorname{Im} z_1\bar{z}_1 \\ w_3 &= (y_1^2 + x_1^2) - (y_2^2 + x_2^2) = z_1\bar{z}_1 - z_2\bar{z}_2 \\ w_4 &= y_1^2 + x_1^2 + y_2^2 + x_2^2 = z_1\bar{z}_1 + z_2\bar{z}_2. \end{aligned}$$

The Hopf variables satisfy the relation

$$w_1^2 + w_2^2 + w_3^2 = w_4^2 \quad \& \quad w_4 \geq 0. \quad (9)$$

Let

$$\Phi : S^1 \times \mathbf{R}^4 \longrightarrow \mathbf{R}^4 : (t, (x, y)) \longmapsto \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

be the $S^1 = \mathbf{R}/2\pi\mathbf{Z}$ action on \mathbf{R}^4 defined by the flow of the harmonic oscillator vectorfield X_{H_2} . Consider the space \mathbf{R}^4/S^1 each point of which is an orbit of X_{H_2} . Using the S^1 -invariants w_i as coordinates,⁴ it follows that the orbit space \mathbf{R}^4/S^1 is defined by (9). Note that \mathbf{R}^4/S^1 is a cone on a 2-sphere and hence is topologically just \mathbf{R}^3 . Indeed \mathbf{R}^4/S^1 is not smooth because O is an equilibrium point of X_{H_2} .

To remove the symmetry of H generated by the flow of the harmonic oscillator, restrict H to the level set $H_2^{-1}(h) = \{\frac{1}{2}(y_1^2 + x_1^2 + y_2^2 + x_2^2) = h\}$. This defines a function

$$H_{2h}(w_1, w_2, w_3) = \mathcal{H}(w_1, w_2, w_3, 2h)$$

on the space $H_2^{-1}(h)/S^1$ of orbits of the harmonic oscillator of energy $2h$. Concretely, $H_2^{-1}(h)/S^1$ is defined by⁵

$$w_1^2 + w_2^2 + w_3^2 = w_4^2 \quad \& \quad w_4 = 2h,$$

that is,

$$w_1^2 + w_2^2 + w_3^2 = (2h)^2,$$

which is the standard 2-sphere S_{2h}^2 of radius $2h$ in \mathbf{R}^3 .

To obtain Hamilton's equations on S_{2h}^2 corresponding to the reduced Hamiltonian H_{2h} we first note that $\{w_i\}_{i=1}^4$, thought of as functions on \mathbf{R}^4 span a Lie algebra \mathcal{L} under Poisson bracket $\{ , \}$ with bracket relations

$$\begin{aligned} w_k &= -4 \sum_{i,j=1}^3 \epsilon_{kij} \{w_i, w_j\}, \quad \text{for } i, j, k \in \{1, 2, 3\} \\ 0 &= \{w_4, w_i\}, \quad \text{for all } i \end{aligned}$$

⁴For a proof, see [5] p. 34.

⁵For a treatment of the regular case of reduction of symmetry in Hamiltonian systems, see [6], p.299. This theory is inadequate to handle cases where the symmetry has fixed points. Our treatment follows the ideas of singular reduction see,[7].

Observe that $(\mathcal{L}, \{ , \})$ is isomorphic as a Lie algebra to $u(2)$. Next, for $j \in \{1, 2, 3\}$ we compute

$$\begin{aligned}\dot{w}_j &= \{w_j, H_{2h}\} = \sum_{k=1}^3 \frac{\partial H_{2h}}{\partial w_k} \{w_k, w_j\} \\ &= -4 \sum_{k,\ell=1}^3 \varepsilon_{jk\ell} \frac{\partial H_{2h}}{\partial w_k} w_\ell \\ &= -4(\nabla H_{2h} \times w)_j.\end{aligned}\tag{10}$$

In other words, Hamilton's equations are a generalization of Euler's equations. Note that these equations are actually defined on \mathbf{R}^3 and that S_{2h}^2 is an invariant manifold.

We summarize what we have done. To remove the harmonic oscillator symmetry of the original invariant Hamiltonian H , we found a reduced phase space S_{2h}^2 , a reduced Hamiltonian H_{2h} and a set of equations which describe the dynamics on the reduced space. The mapping

$$\Pi : H_2^{-1}(h) \subseteq \mathbf{R}^4 \longrightarrow S_{2h}^2 \subseteq \mathbf{R}^3 : (x, y) \longrightarrow (w_1, w_2, w_3),$$

which performs this symmetry reduction, is the Hopf mapping.

5. HENON-HEILES HAMILTONIAN

In this section we apply the foregoing normal form theory to study the Henon-Heiles Hamiltonian

$$H(x, y) = \frac{1}{2}(y_1^2 + y_2^2) + \frac{1}{2}(x_1^2 + x_2^2) + \frac{1}{3}x_1^3 - x_1x_2^2$$

on \mathbf{R}^4 with canonical coordinates (x, y) .

First we make some general remarks about the Henon-Heiles system.⁶

1). The Hamiltonian H is invariant under the \mathbf{Z}_3 action on \mathbf{R}^4 generated by the matrix

$$\begin{pmatrix} R & 0 \\ 0 & R \end{pmatrix} \quad \text{where} \quad R = \begin{pmatrix} \cos \frac{2\pi}{3} & -\sin \frac{2\pi}{3} \\ \sin \frac{2\pi}{3} & \cos \frac{2\pi}{3} \end{pmatrix}.$$

2). For $0 < h < \frac{1}{6}$ the image of the energy surface $H^{-1}(h)$ under the projection mapping

$$\pi : \mathbf{R}^4 \longrightarrow \mathbf{R} : (x, y) \longrightarrow x$$

has four connected components, called Hill regions. The motion takes place in these regions. We are only interested in motions in the Hill region containing the origin (see Figure 4).

3). For every $0 < h < \frac{1}{6}$, the Henon-Heiles vectorfield X_H has *eight* fundamental periodic solutions of short period. They are shown in Figure 5 with their linear stability type indicated by an “e” if elliptic and “h” if hyperbolic.⁷

4). A Poincaré section S_h for X_H is constructed as follows. Set $x_2 = 0$ in H obtaining

$$h = \frac{1}{2}(y_1^2 + y_2^2) + \frac{1}{2}x_1^2 + \frac{1}{3}x_1^3.$$

⁶The original treatment of the Henon-Heiles Hamiltonian is [8]. A good survey of rigorously proved facts about the Henon-Heiles Hamiltonian can be found in [9].

⁷Existence of the fundamental periodic solutions is proved in [2], p. 411.

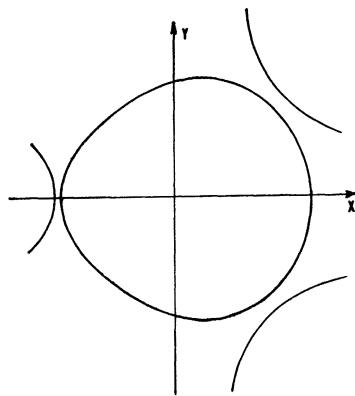


Figure 4. The Hill regions for the Hénon-Heiles system.

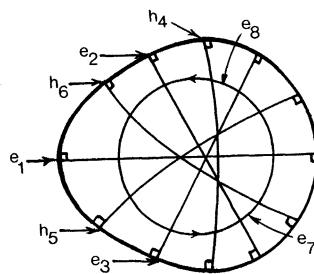


Figure 5. The fundamental periodic solutions of the Hénon-Heiles system.

Since $y_2^2 \geq 0$, for small values of h we find that \mathcal{S}_h , defined by

$$h \geq \frac{1}{2}y_1^2 + \frac{1}{2}x_1^2 + \frac{1}{3}x_1^3 \quad \& \quad x_2 = 0 \quad \& \quad y_2 = \sqrt{2h - (y_1^2 + x_1^2 + \frac{2}{3}x_1^3)},$$

is an oval lying in the energy surface $H^{-1}(h)$ and is transverse to $X_H(x, y)$ at each $(x, y) \in \mathcal{S}_h$. Therefore \mathcal{S}_h is a Poincaré section for X_H . It is *not* a global cross section because it is not transverse to every orbit of X_H . Numerically computing the Poincaré mapping defined by successive intersections of the integral curves of X_H with \mathcal{S}_h gives the following pictures (Figure 6).⁸

The figure on the left seems to show that the orbits of the Poincaré map lie on the level sets of some function when h is small, say $\sim \frac{1}{24}$. Because \mathcal{S}_h is *not* a global cross section, the number of fixed points (indicated by large dark points in Figure 6a) of the Poincaré map, which correspond to fundamental periodic orbits of X_H , do *not* add up to eight. On the other hand when h is large, say $\sim \frac{1}{8}$ the orbits look randomly scattered in the figure on the right.

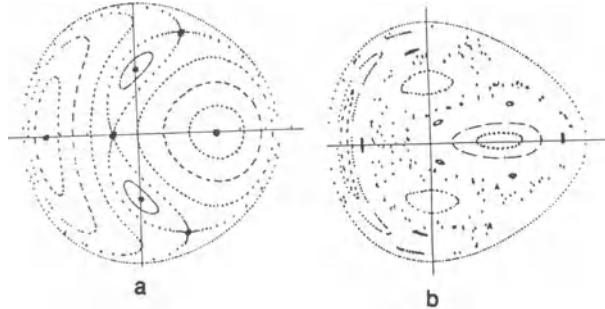


Figure 6. Poincaré maps for the Hénon-Heiles system for different energy values (determined by the constraint $h = E$): (a) $E = 1/24$, (b) $E = 1/8$.

We now try to explain this low energy phenomenon using normal form and reduction. The normal form in Hopf variable of the Hénon-Heiles Hamiltonian up to sixth order terms⁹ is

$$\tilde{H} = H_2 + \varepsilon^2 H_4 + \varepsilon^4 H_6,$$

where

$$\begin{aligned} H_2 &= \frac{1}{2}w_4 \\ H_4 &= \frac{1}{48}(7w_2^2 - 5w_4^2) \\ H_6 &= \frac{1}{64}\left(-\frac{67}{54}w_4^3 - \frac{7}{18}w_2^2w_4 - \frac{28}{9}w_3^3 + \frac{28}{3}w_1^2w_3\right). \end{aligned}$$

⁸These pictures are taken from [8].

⁹The first computation of the normal form to sixth order was done in [10]. A more modern calculation using Lie series is given in [2].

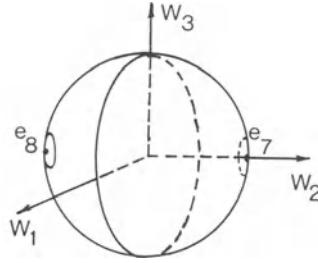


Figure 7. Elliptic periodic orbits and critical circle of the first-order normalized reduced Hamiltonian.

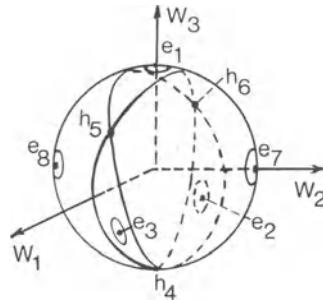


Figure 8. Elliptic and hyperbolic orbits of the second-order normalized reduced Hamiltonian.

The reduced normalized Hamiltonian on S_{2h}^2 to first order is

$$H_{2h}^1 = \frac{1}{48}(7w_2^2 - 5(2h)^2)$$

and to second order is

$$H_{2h}^2 = \frac{1}{64}\left(-\frac{67}{54}(2h)^3 - \frac{7}{18}(2h)w_2^2 - \frac{28}{9}w_3^3 + \frac{28}{3}w_1^2w_3\right).$$

We now analyze H_{2h}^1 geometrically. On S_{2h}^2 critical points of H_{2h}^1 occur when the vector

$$\nabla H_{2h}^1 = \left(0, \frac{7}{24}w_2, 0\right)$$

is normal to S_{2h}^2 . This happens at two points $\pm(0, 2h, 0)$ and at the circle $\{w_2 = 0\} \cap S_{2h}^2$. The critical points are elliptic and correspond to elliptic periodic orbits of the Hamiltonian vectorfield $X_{H_{2h}^1}$ of the first order normalized reduced Hamiltonian. These periodic orbits persist to the periodic orbits e_7 and e_8 of the full Henon-Heiles vectorfield X_H .

The critical circle corresponds to an *unresolved* nonlinear resonance (see Figure 7). To analyze this resonance we look at the second order reduced normalized Hamiltonian H_{2h}^2 . Then

$$\nabla H_{2h}^2 = \frac{1}{64}\frac{28}{3}(2w_1w_3, (2 - \frac{1}{6}h)w_2, w_1^2 - w_3^2)$$

is normal to S_{2h}^2 at the points,

$$\begin{aligned} e_{7,8} &= \pm(0, 2h, 0) \\ e_1 &= (0, 0, 2h) & e_2 &= Se_1 & e_3 &= Se_2 \\ h_4 &= (0, 0, -2h) & h_5 &= Sh_4 & h_6 &= Sh_5, \end{aligned}$$

where $S = \begin{pmatrix} \cos \frac{4\pi}{3} & 0 & \sin \frac{4\pi}{3} \\ 0 & 1 & 0 \\ -\sin \frac{4\pi}{3} & 0 & \cos \frac{4\pi}{3} \end{pmatrix}$, which comes from the Z_3 symmetry of the Henon-Heiles Hamiltonian. We see that the second order normal form has resolved the nonlinear resonance (see Figure 8). One can show that these six new periodic solutions $e_1, e_2, e_3, h_4, h_5, h_6$ persist to the full Henon-Heiles Hamiltonian and that they correspond to the remaining fundamental periodic solutions.¹⁰

Actually we have done more than this. The flow of the reduced normalized Hamiltonian gives a nondegenerate twist map approximation to the flow of the original Henon-Heiles Hamiltonian in a neighborhood of each elliptic critical point. Even though the Henon-Heiles vectorfield is nonintegrable,¹¹ we are in position to apply KAM theory to prove the nonlinear stability of these elliptic periodic orbits.

REFERENCES

- [1] R. Cushman and D. Rod, 1982, “Reduction of the semisimple 1:1 resonance”, *Physica D*, **6**, 105-112.
- [2] R. Churchill, M. Kummer, and D. Rod, 1983, “On averaging, reduction and symmetry in Hamiltonian systems”, *J. Diff. Eqns.*, **49**, 359-414.
- [3] R. Cushman, 1992, “A survey of normalization techniques applied to perturbed Keplerian systems”, *Dynamics reported*, (new series) **1**, 54-112, (New York: Springer).
- [4] J-C. van der Meer, 1985, *The Hamiltonian Hopf bifurcation*, Lecture Notes in Mathematics, **1185**, (New York: Springer).
- [5] R. Cushman and L. Bates, *Global Aspects of Classical Integrable Systems*, (to appear).
- [6] R. Abraham and J. Marsden, 1978, *Foundations of Mechanics, second edition*, (Reading, Mass.: Benjamin-Cummings).
- [7] J. Arms, R. Cushman, and M. Gotay, 1991, “A universal reduction procedure for Hamiltonian group actions”, in: *The Geometry of Hamiltonian Systems*, ed. T. Ratiu, 31-51, (Boston: Birkhauser).
- [8] M. Henon and C. Heiles, 1964, “The applicability of the third integral of motion: some numerical experiments”, *Astronomical Journal*, **69**, 73-79.
- [9] R. Churchill, G. Pecelli, and D. Rod, 1978, “A survey of the Henon-Heiles Hamiltonian with applications to related examples”, in: *Como Conference Proceedings on Stochastic Behavior in Classical and Quantum Hamiltonian Systems*, eds. G. Casati and J. Ford, 76-136, Springer Lecture Notes in Physics, **93**, (New York: Springer Verlag).
- [10] F. Gustavson, 1966, “On constructing formal integrals of a Hamiltonian system near an equilibrium point”, *Astronomical Journal*, **71**, 670-686.
- [11] R. Churchill and D. Rod, 1980, “Pathology in dynamical systems III. Analytic Hamiltonians”, *J. Diff. Eqns.*, **37**, 23-38.

¹⁰See [2] p. 411.

¹¹See [11] p. 35 for a proof.

BETWEEN INTEGRABILITY AND CHAOS

David L. Rod† and Richard C. Churchill‡

†Dept. of Mathematics and Statistics
University of Calgary, Calgary, Alberta, T2N 1N4 Canada

‡Dept. of Mathematics, Hunter College and Graduate School, CUNY
695 Park Avenue, New York, New York, 10021, USA

Abstract. We describe a method for proving the non-integrability of an analytic Hamiltonian system which does not appeal to the existence of chaotic motion.

1. INTRODUCTION

An important consequence of chaos in a given Hamiltonian system is non-integrability [4]. While methods for establishing chaotic behavior are treated elsewhere in the volume, here the purpose is to inform readers of a recent technique, which has now been significantly refined, for proving non-integrability without consideration of chaotic behavior, although such behavior is not precluded. The ideas have been used, for example, to establish the non-integrability of the classical Zeeman Hamiltonian governing the three-dimensional magnetized hydrogen atom [3]. Applications to complexity in the spatio-temporal dynamics of certain biological models can be found in [6].

2. NON-INTEGRABILITY OF HAMILTONIAN SYSTEMS

When employing the method the Hamiltonian under study must be complex analytic or meromorphic and time must be complex. When successfully applied to a complexified real analytic Hamiltonian the consequence is that the real system can have no real entire (e.g. polynomial) analytic integral independent of the Hamiltonian, nor an independent

integral which is a quotient of real entire functions (e.g., a rational integral). (“Entire” means: infinite radius of convergence.)

The ideas are most easily conveyed in terms of a two-degree of freedom system on $\mathbb{C}^4 \approx \{(x, x_2, y, y_2)\}$ assuming the symplectic structure $\omega = dx \wedge dy + dx_2 \wedge dy_2$. Specifically, assume $H : \mathbb{C}^4 \rightarrow \mathbb{C}$ is a meromorphic Hamiltonian of the form

$$H(x, x_2, y, y_2) = A(x, y) + (1/2)B(x, y)x_2^2 + C(x, y)x_2y_2 \\ + (1/2)D(x, y)y_2^2 + \mathcal{O}_3(x_2, y_2),$$

and note that the corresponding vector field X_H is tangent to the (x, y) -plane where defined.

Now suppose Γ is a maximal non equilibrium phase curve of X_H , i.e., the image, after maximal analytic continuation, of a local solution of the corresponding Hamiltonian system. Assume that Γ is contained in the xy -plane so as to insure that the variational (“linearized”) equations along this curve decouple as

$$\begin{aligned}\dot{r} &= A_{xy}r + A_{yy}s \\ \dot{s} &= -A_{xx}r - A_{xy}s\end{aligned}$$

and

$$\begin{aligned}\dot{u} &= C(x, y)u + D(x, y)v \\ \dot{v} &= -B(x, y)u - C(x, y)v.\end{aligned}$$

The first equation is none other than the variational equation along Γ when considered as a phase curve of the planar Hamiltonian field $X_{\hat{H}}$, where $\hat{H} : (x, y) \mapsto H(x, 0, y, 0)$ and the symplectic structure is $\hat{\omega} = dx \wedge dy$.

The second equation is the crucial one for our purposes: this *normal variational equation* (NVE) of X_H along Γ is to be considered a linear differential equation on this Riemann surface (with local parameter provided by t). In differential geometric terms it can be regarded as a global frame representation of a holomorphic connection ∇ on the symplectic *normal bundle* $E := (T\Sigma|\Gamma)/T\Gamma$ over Γ , where Σ is the energy surface containing Γ . This more abstract interpretation does not require any special form for H , nor the two degree-of-freedom restriction. Indeed, one can work on any $2n$ -dimensional complex analytic symplectic manifold. (In the more general situation the definitions of the normal bundle and the NVE are modified when additional H -independent integrals are known explicitly.)

3. ZIGLIN'S THEOREM

Now consider the monodromy representation of the NVE. In terms of the second equation this can be viewed as the homomorphism $\rho : \pi_1(\Gamma) \rightarrow SL(2, \mathbb{C})$ defined by analytically continuing the fundamental matrix solution around inverses of loops in Γ based at some distinguished point p (so $\pi_1(\Gamma) := \pi_1(\Gamma, p)$). In the more general setting we have $\rho : \pi_1(\Gamma, p) \rightarrow Sp(E_p)$, where $Sp(E_p)$ is the group of symplectic automorphisms of the fibre E_p of E over p . From either perspective the *monodromy group* $G := \rho(\pi_1(\Gamma, p))$ acts as a group of linear operators on a vector space V ; in the motivating case one simply deals with a group of matrices. A rational function f on V is an *invariant* of this action

if $f(g \cdot v) = f(v)$ for all $g \in G$ and $v \in V$, e.g., the polynomial $z = (z_1, z_2) \mapsto z_1 z_2$ on \mathbb{C}^2 is an invariant of the action of the diagonal subgroup of $SL(2, \mathbb{C})$.

Theorem [7]: Suppose M is a $2n$ -dimensional complex symplectic manifold, X_H is a meromorphic Hamiltonian vector field on M , and $G \subset Sp(E_p)$ is the monodromy group of the NVE of some phase curve Γ of X_H . Then each meromorphic integral of X_H defined on some neighborhood of Γ and functionally independent of H induces a rational invariant of G , and these induced invariants can be assumed algebraically independent over \mathbb{C} when the integrals are functionally independent.

In the statement functional independence of the integrals is assumed to hold on some neighborhood of Γ , but not necessarily on the set itself. Indeed, individual meromorphic integrals are permitted to “blow up” on this Riemann surface.

For the two-degree of freedom example under consideration the consequence is: if X_H admits an H -independent integral the matrix group $G \subset SL(2, \mathbb{C})$ must have a rational invariant. By proving there is no such invariant one proves there is no such integral.

The original proof of the theorem is found in [7]; for additional detail see [1], which is self-contained. For a proof tailored to the context of our \mathbb{C}^4 -example see [2]; for a worked out example in this setting see [2] or [5].

When the NVE is Fuchsian one can replace the monodromy group in the statement of the theorem with the differential Galois group, which is far easier to compute. See [1] for details, including the necessary background from algebraic groups and differential Galois theory.

REFERENCES

- [1] A. Baider, R. C. Churchill, D. L. Rod and M. Singer, “On the Infinitesimal Geometry of Integrable Systems,” preprint (1992), The Fields Institute, Waterloo, Ontario.
- [2] H. Ito, “Non-Integrability of Hénon-Heiles and a Theorem of Ziglin,” *Kodai Math. J.*, **8** (1985), pp. 120-138.
- [3] M. Kummer and A. W. Sáenz, “Non-integrability of the Classical Zeeman Hamiltonian”, preprint (1993), University of Toledo.
- [4] J. Moser, “Stable and Random Motions in Dynamical Systems”, *Ann. of Math. Studies* **77**, Princeton University Press, Princeton, 1973.
- [5] D. L. Rod, “On a Theorem of Ziglin in Hamiltonian Dynamics, in Hamiltonian Dynamical Systems”, (Kenneth R. Meyer and Donald E. Saari, Eds.,) *Contemporary Mathematics*, **81**, Am. Math. Soc., Providence, 1988.
- [6] D. L. Rod and B. Sleeman, “Complexity in Spatio-Temporal Dynamics”, to appear.
- [7] S. L. Ziglin, “Branching of Solutions and Non-existence of First Integrals in Hamiltonian Mechanics I and II”, *Functional Anal. Appl.*, **16** (1982), pp. 181-189, and **17** (1983), pp. 6-17.

ON DEFINING CHAOS IN THE ABSENCE OF TIME

Richard C. Churchill

Dept. of Mathematics, Hunter College and Graduate School, CUNY
695 Park Avenue, New York, New York, 10021, USA

Abstract. One of the standard definitions of a chaotic dynamical system on a metric space involves three conditions, one being sensitive dependence on the choice of initial values. Using the recent discovery that the sensitivity hypothesis is a logical consequence of the other two conditions we formulate a time-and-metric independent concept of chaos for foliations which implies the usual definition when the leaves are the orbits of a flow on a manifold. Simple examples are presented to make the point that any reference to “chaos” when either or both of the other standard conditions has not been verified could be quite misleading. In particular, for any integer $n > 1$ we give an example of a completely integrable n -degree of freedom Hamiltonian system, with compact energy surfaces, having the property that the induced flows on almost all energy surfaces admit sensitive dependence on initial conditions.

1. INTRODUCTION

The definition of “chaos” in nonlinear dynamical systems as an extreme sensitivity of the solutions upon the given initial conditions generally requires that one compute the time evolution of neighboring trajectories from infinitesimally separated initial conditions over a time period long enough that the time averaged convergence or divergence of the trajectories stabilizes to a measurable becomes value. This process assumes that the time coordinate used to parametrize the dynamical has an invariant meaning. In the case of general relativity, this is not true and this fact has caused problems in measuring the chaotic behavior of the Mixmaster cosmologies. The intent of this paper is to define “chaos” in a manner that is independent of a time parameter and therefore may be applicable to problems in general relativity.

2. CONDITIONS FOR DEFINING “CHAOS”

A vector field on a smooth manifold defines a foliation away from the collection of critical points; the orbits of the associated flow are the leaves, and for ease of exposition we also refer to critical points as leaves (a “foliation with degeneracies”). More generally, suppose \mathcal{F} is a foliation of a smooth manifold M by smooth submanifolds (degeneracies allowed). \mathcal{F} is *chaotic* if:

- (a) the compact leaves of \mathcal{F} are dense in M and there are at least two distinct compact leaves; and
- (b) \mathcal{F} is *transitive*, i.e., given any two non empty open sets $U, V \subset M$ there is an $m \in U$ such that the leaf through m intersects V non trivially.

Because a vector field specifies an orientation on non trivial orbits one might include an orientability hypothesis in this definition. For example, in the case of a foliation by curves one could replace (b) by

- (b') \mathcal{F} is *positively transitive*, i.e., given any two non empty open sets $U, V \subset M$ there is an $m \in U$ such that the positive leaf through m intersects V non trivially.

Here time enters the picture, but only in the sense that past and future are now distinguished.

Smoothness is not essential in these definitions: they can be formulated on any topological space provided one is precise about the meanings of “foliation” and “oriented leaves.” We leave this to the interested reader; for curves the meaning is clear.

To motivate the two definitions we need a few preliminaries.

Throughout X is a topological space and $\varphi: \mathbb{R} \times X \rightarrow X$ is a continuous flow. (\mathbb{R} denotes the real field with the usual topology.) Write $\varphi(t, x)$ as $t \cdot x$, and $\varphi(S, Y)$, for any $\emptyset \neq S \subset \mathbb{R}$ and any $\emptyset \neq Y \subset X$, as $S \cdot Y$. For any $x \in X$ denote the φ -orbit $\{t \cdot x : t \in \mathbb{R}\}$ by $\mathcal{O}(x)$.

Lemma: *When X is metrizable and $x \in X$ the following statements are equivalent:*

- (a) $\mathcal{O}(x)$ is compact; and
- (b) $\mathcal{O}(x)$ is periodic (possibly a rest point).

Proof: Assume (a) and write $\mathcal{O}(x)$ as $\bigcup_{n=1}^{\infty} [-n, n] \cdot x$. Then by Baire some $[-n, n] \cdot x$ contains an open (rel $\mathcal{O}(x)$) set U , and the fact that we can choose a finite subcover of the open cover $\{t \cdot U\}_{t \in \mathbb{R}}$ of $\mathcal{O}(x)$ easily implies that $\mathcal{O}(x) \subset [-m, m] \cdot x$ for some $m > 0$. But then $2m \cdot x = s \cdot x$ for some $s \in [-m, m]$, i.e., $(2m - s) \cdot x = x$, and periodicity follows.

The reverse implication is obvious. Q.E.D.

Now suppose X is a metric space (with metric d). We say that φ has *sensitive dependence on initial conditions*, or simply that φ is δ -*sensitive*, if there is a constant $\delta > 0$ with the following property: for all $x \in X$ and all $\varepsilon > 0$ there is a $y \in X$ satisfying $d(x, y) < \varepsilon$ and $d(t \cdot x, t \cdot y) > \delta$ for some $t > 0$.

Again assume X is a metric space. The flow φ is *chaotic* if:

- (i) the periodic points are dense and there are at least two distinct periodic orbits;
- (ii) φ is transitive, i.e., for any two non empty open subsets $U, V \subset X$ there is a $t > 0$ such that $(t \cdot U) \cap V \neq \emptyset$; and

- (iii) φ is δ -sensitive for some $\delta > 0$, i.e., there is sensitive dependence on initial conditions.

The definition is from [2]; others are also encountered.

The geodesic flow on the unit tangent bundle of a compact manifold of negative curvature is the prototypical chaotic (and prototypical ergodic) dynamical system.

It was noted recently (see [1]) that Condition (iii) is a logical consequence of (i) and (ii), and can therefore be dropped from the definition. This is our justification for the two definitions of a chaotic foliation given at the beginning of the paper. Indeed, in view of the Lemma we have simply reformulated (i) and (ii) in the context of foliations. In the proposition below we offer a mild generalization of the (i) & (ii) \Rightarrow (iii) result.

We adopt the following notation: for any $x \in X$ and any constant $\lambda > 0$ we let $B_\lambda(x) := \{y \in X : d(x, y) < \lambda\}$; for any non empty $K \subset X$ we let $d(x, K) := \inf\{d(x, k) : k \in K\}$; for any non empty $L \subset X$ we let $d(K, L) := \inf\{d(k, \ell) : k \in K, \ell \in L\}$.

Recall that a subset $M \subset X$ is *minimal* (w.r.t. φ) if M is non empty, closed, φ -invariant, and has no proper subset with these properties. Examples of minimal sets are: rest points; periodic orbits; tori with irrational flows.

Proposition: Suppose there is a constant $\delta > 0$ and a collection $\mathcal{M} = \{M_\beta\}$ of compact minimal sets in X such that:

1. $\bigcup_\beta M_\beta$ is dense in X ;
2. for any β and any $q \in M_\beta$ there is a non empty φ -invariant set $N_{\beta,q} \subset X$ with the following two properties:
 - (a) $d(q, N_{\beta,q}) > 4\delta$; and
 - (b) $M_\beta \cup N_{\beta,q}$ is contained within some transitive φ -invariant subset of X .

Then φ is δ -sensitive.

The proof is a minor modification of the argument in [1].

Proof: Given $p \in X$ and $\varepsilon > 0$ choose:

$q \in B_\varepsilon(p)$ such that $q \in M_\beta$ for some $M_\beta \in \mathcal{M}$ (which is possible by 1);

$T > 0$ so large that any interval $[t, t+T] \subset \mathbb{R}$ contains a point s such that $d(s \cdot q, q) < \delta$ (this is where minimality is needed; see p. 378 of [3]);

$m' \in N_{\beta,q}$ at random;

an open neighborhood V of m' such that $d(t \cdot v, t \cdot m') < \delta$ for all $v \in V$ and all $0 \leq t \leq T$ (the existence of such a V follows easily from the compactness of $[0, T]$);

$(t_r, r) \in (0, \infty) \times B_\varepsilon(p)$ such that $r' := t_r \cdot r \in V$ (which is possible by 2(b)); and

$s \in [t_r, t_r + T]$ such that $d(s \cdot q, q) < \delta$ (which is possible by the choice of T).

Now set $m := -t_r \cdot m' \in N_{\beta,q}$. We then have $s \cdot r = (s - t_r) \cdot r'$, $s \cdot m = (s - t_r) \cdot m'$, and because $0 \leq s - t_r \leq T$ it follows from the choice of V that

$$d(s \cdot r, s \cdot m) < \delta.$$

To complete the proof simply note (see 2(a)) that

$$4\delta < d(q, s \cdot m) \leq d(q, s \cdot q) + d(s \cdot q, s \cdot m)$$

implies

$$\begin{aligned} 3\delta &< d(s \cdot q, s \cdot m) \leq d(s \cdot q, s \cdot p) + d(s \cdot p, s \cdot m) \\ &\leq d(s \cdot q, s \cdot p) + d(s \cdot p, s \cdot r) + d(s \cdot r, s \cdot m), \end{aligned}$$

which in turn implies

$$2\delta < d(s \cdot p, s \cdot q) + d(s \cdot p, s \cdot r).$$

But then $d(s \cdot p, s \cdot q) > \delta$ or $d(s \cdot p, s \cdot r) > \delta$ (or both). Q.E.D.

Corollary 1: Suppose the compact minimal sets are dense, there are at least two such sets, and φ is transitive. Then φ has sensitive dependence on initial conditions.

Proof: Choose distinct compact minimal sets M_1, M_2 and let $10\delta = d(M_1, M_2)$; since $M_1 \cap M_2$ is φ -invariant and the M_j are minimal the two sets must be disjoint, and therefore $\delta > 0$. Using the triangle inequality one now sees easily that any $x \in X$ has distance greater than 4δ either from M_1 or M_2 (or both). In the statement of the proposition we can therefore choose $N_{\beta, q} := M_1$ or $N_{\beta, q} := M_2$. Q.E.D.

Corollary 2 [1]: Suppose the periodic orbits are dense, there are at least two such orbits, and φ is transitive. Then φ has sensitive dependence on initial conditions.

In other words: (i) and (ii) imply (iii).

Proof: Periodic orbits are compact minimal sets. Q.E.D.

As will be seen from the following examples, Corollary 2 gives the only dependency relation between (i), (ii) and (iii).

3. EXAMPLES

1. A rational flow on a 2-torus satisfies (i) but not (ii) and not (iii).
2. An irrational flow on a 2-torus satisfies (ii) but not (i) and not (iii).
3. The flow of the vector field $(x_2^2 + 1)\frac{\partial}{\partial x_1}$ on \mathbb{R}^2 satisfies (iii) but not (i) and not (ii).
4. Let X be vector field on a 2-torus T^2 governing an irrational flow, and let $g : T^2 \rightarrow [0, 1]$ be a smooth function which vanishes at precisely one point $p \in T^2$ and is identically 1 off some small disc centered at p . Then the flow of gX satisfies (ii) and (iii) but not (i).
5. Suppose $M \subset \mathbb{R}^3$ is a level set of some smooth function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ and is regular in the sense that $\nabla f(x) \neq 0$ for each $x \in M$; then for each $x \in M$ we can define the normal vector $n(x)$ to M at x as $|\nabla f(x)|^{-1}\nabla f(x)$. Now suppose $U \subset \mathbb{R}^3$ is an open neighborhood of M where $\nabla f(x)$ does not vanish, let $H : U \rightarrow \mathbb{R}$ be any smooth function, and consider the differential equation

$$\dot{x} = \nabla H(x) \times n(x)$$

on U , where \times denotes the usual cross product on \mathbb{R}^3 . One checks easily that f and H are first integrals (i.e., constants of the motion), and as a result the equation induces a flow on M having each orbit contained within the intersection of a level set of H with this surface.

It is worth noting that the equation of the previous paragraph can be viewed as a one degree-of-freedom (and therefore completely integrable) Hamiltonian system

on M . Indeed, a symplectic structure ω on this surface is defined by $\omega(x)(u, v) := \langle n(x), u \times v \rangle$, where $\langle \cdot, \cdot \rangle$ is the usual inner product, and our assertion then follows from the observation that for the vector field X_H on M defined by the equation we have $i_{X_H}\omega = dH$ ($:= d(H|M)$).

Now consider the case

$$f : x = (x_1, x_2, x_3) \mapsto 4(x_1^2 + x_2^2) - ((15/16) + |x|^2)^2,$$

$M := f^{-1}(\{0\})$, $U = \mathbb{R}^3$, and $H : x \mapsto (1/2)|x|^2$. Here we have $\nabla H(x) = x$, and the equation $\dot{x} = \nabla H(x) \times n(x)$ reduces to

$$\dot{x} = x \times n(x).$$

M is the 2-torus obtained by rotating the circle $\{(x_1, x_3) : (x_1 - 1)^2 + x_3^2 = 1/16\}$ about the x_3 -axis, the level sets of H (away from the origin) are spheres, and the non empty intersections of these level sets with M consist of (one or two) circles parallel to (or within) the x_1x_2 -plane. It is a simple matter to check that the particular circles contained in the x_1x_2 -plane are made up of fixed points, that all other circles are periodic orbits, and that the (signed) period of these orbits may be viewed locally as a non-constant analytic function of $x_1^2 + x_2^2$ which reaches a (local) extremum at $x_1^2 + x_2^2 = (31/32)^2$ and is strictly monotonic elsewhere. It follows easily that the flow on M satisfies (i) and (iii) but not (ii).

Using Example 5 we now construct, for any integer $n > 1$, a completely integrable n -degree of freedom Hamiltonian vector field on a compact symplectic manifold with the property that the induced flows on almost all positive energy surfaces have sensitive dependence on initial conditions. As previously stated this points out the dangers of concluding, simply from sensitive dependence, that one is dealing with a chaotic system, *even in the case of compact energy surfaces*.

To describe the example write $x \in \mathbb{R}^{3n}$ as $x = (x_1, \dots, x_n)$, where $x_j = (x_{j1}, x_{j2}, x_{j3}) \in \mathbb{R}^3$, $j = 1, \dots, n$, and consider the differential equation

$$\dot{x}_j = x_j \times n(x_j), \quad j = 1, \dots, n,$$

where $n = n(x_j) = |\nabla f(x_j)|^{-1} \nabla f(x_j)$ is constructed from the function f introduced in the final paragraph of Example 5. This system is Hamiltonian on the n -fold Cartesian product $\tilde{M} := M \times \cdots \times M \subset \mathbb{R}^{3n}$, where M is the torus introduced in Example 5, relative the the symplectic structure ω defined by

$$\omega(x)(u, v) = \sum_j \langle n(x_j), u_j \times v_j \rangle;$$

the Hamiltonian is $H : x \in \tilde{M} \mapsto (1/2)\sum_j |x_j|^2 = (1/2)|x|^2$. The system is, moreover, completely integrable: the functions $H_j : x \in \tilde{M} \mapsto (1/2)|x_j|^2$ provide n (independent) integrals in involution.

Fix $h > 0$ and let $\Sigma_h := H^{-1}(\{h\}) \subset \tilde{M}$ denote the (compact) energy surface of energy h . If $p = (p_1, \dots, p_n) \in \Sigma_h$ then for almost all h for which $\Sigma_h \neq \emptyset$ there is a $q = (q_1, \dots, q_n) \in \Sigma_h$ arbitrarily close to p such that p_1 and q_1 lie on different circles of M (as described in Example 5), and the φ -orbits of p and q therefore project to distinct periodic orbits on (the first copy of) M (within $\tilde{M} = M \times \cdots \times M$). Since these projections on M have distinct periods one sees easily that $d(s \cdot p, s \cdot q) \geq 3/2$ (= the

diameter of the center hole of the torus M) for some $s > 0$, and sensitivity in the initial conditions follows easily (e.g., take $\delta = 1$).

In closing we note, in view of the Proposition, that Condition (a) in the two definitions of a chaotic foliation \mathcal{F} can be generalized by:

(a') the compact minimal sets of \mathcal{F} are dense in M and there are at least two distinct compact minimal sets.

Here “minimal set” means: a non empty closed union of leaves having no proper subset with these properties.

ACKNOWLEDGEMENTS

My inspiration for writing this note came directly from the conference, and in particular from conversations with Jędrzej Sniatycki. It was Jędrzej who suggested formulating a notion of chaos in terms of foliations.

REFERENCES

- [1] J. Banks, J. Brooks, G. Cairns, G. Davis and P. Stacey, “On Devaney’s definition of Chaos”, *Am. Math. Monthly* **99** (1991), pp. 332-334.
- [2] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd-ed., Addison-Wesley, Redwood City, CA., 1989.
- [3] V.V. Nemytskii and V.V. Stepanov, *Qualitative Theory of Ordinary Differential Equations*, Princeton Univ. Press, Princeton, 1960.

ON THE DYNAMICS OF GENERATORS OF CAUCHY HORIZONS

P. T. Chruściel[†]¹ and J. Isenberg[‡]²

[†]Max Planck Institut für Astrophysik,
D 85740 Garching bei München, Germany

[‡]Department of Mathematics and Institute for Theoretical Science,
University of Oregon, Eugene, OR 97403, USA

Abstract. We discuss various features of the dynamical system determined by the flow of null geodesic generators of Cauchy horizons. Several examples with non-trivial global behaviour are constructed.

1. INTRODUCTION

In considering the question of whether the laws of physics prevent one from constructing a “time machine”, Hawking [1] and Thorne [2] have both stressed the importance of understanding the generic behavior of the null generators of compactly generated Cauchy horizons. In particular it has been suggested (*cf. e.g.* [2, 3] and references therein) that the onset of quantum instabilities in Cauchy horizons containing “fountains” would prevent the formation of time machines. Here a “fountain” on a future Cauchy horizon is defined as a periodic³ generator γ of the horizon such that a “nonzero-measure” set of generators of the horizon asymptotically approaches γ when followed backwards in time. It is therefore of some interest to enquire whether or not the existence of fountains is a generic property of “compactly generated” Cauchy horizons. In this work we wish to point out that this is unlikely to be true: we construct spacetimes with compactly generated Cauchy horizons for which no fountains occur.

¹Alexander von Humboldt fellow, on leave of absence from the Institute of Mathematics, Polish Academy of Sciences, Warsaw. e-mail: piotr@mpa-garching.mpg.de

²e-mail: jim@newton.uoregon.edu

³Throughout this paper “periodic” means periodic as a path in spacetime (and *not* necessarily as a path in the tangent bundle).

When discussing features of Cauchy horizons, one should focus on features which are stable in an appropriate sense. We show that in the set of all spacetimes with compactly generated Cauchy horizons, there are open sets consisting entirely of spacetimes with nonfountain-like behavior. Unfortunately we are able to make rigorous claims only for compact Cauchy horizons. So the possibility remains open that for spacetimes with compactly generated Cauchy horizons which are *not* compact, fountains could generically occur. While it is clear to us that this is not true, we note that there is an important technical difference between compact Cauchy horizons and noncompact yet compactly generated Cauchy horizons: As we show in Section 4, if a Cauchy horizon is compactly generated and noncompact, and if further it is contained in an asymptotically flat spacetime (in a technical sense made precise in that Section), then the generators of the Cauchy horizon cannot be continuous. This is one of the difficulties which one has to face when trying to make any rigorous claims about the dynamics of the generators of some non-compact Cauchy horizons.

It is important to note that, following [1, 2], we do not impose any field equations on the spacetimes under consideration. Recall that one expects the existence of a Cauchy horizon to be an unstable feature, when the Einstein field equations (vacuum, or with energy conditions on the source fields) are imposed. It would be interesting to carefully investigate the extent to which the imposition of field equations restricts the allowed dynamics of Cauchy horizon generators; however this problem is not addressed here.

After discussing some preliminary definitions and ideas in Section 2, we focus on verifying the existence of the spacetimes with nonfountain-like dynamics, first for compact Cauchy horizons (Section 3) and then for compactly generated but noncompact Cauchy horizons (Section 4). The discussion of noncompact Cauchy horizons in Section 4 includes the proof that if the spacetime containing it is asymptotically flat, then the generators cannot be continuous.

2. PRELIMINARIES ON CAUCHY HORIZONS AND DYNAMICAL SYSTEMS

We shall consider C^k , ($k \geq 3$) spacetimes (M^4, g) which contain Cauchy horizons (we use the terminology of [4]). Standard results [4] show that a Cauchy horizon is foliated by a congruence of null geodesics. These are called the *generators* of the horizon. One finds that if one follows a generator of a future Cauchy horizon into its past then the generator always remains inside the horizon. This is not necessarily true if one follows a generator (on a future Cauchy horizon) into its future. We shall say that a future Cauchy horizon \mathcal{H}^+ is *compactly generated*, if there exists a compact set $\mathcal{K} \subset M$ such that every generator of \mathcal{H}^+ enters and remains in \mathcal{K} , when followed into the past.

To discuss the behavior of the generators of a Cauchy horizon, we wish to use some of the language of dynamical systems theory. Recall that a dynamical system (Σ^n, X) consists of an n -dimensional manifold Σ^n and a vector field X specified on Σ^n . Note that a Cauchy horizon \mathcal{H} together with the vector field T of tangents to its generators (normalized in an arbitrary way) constitutes a dynamical system (\mathcal{H}, T) . We shall always choose the *past directed* orientation of the generators on a *future* Cauchy horizon. For future Cauchy horizons the past-oriented generators of \mathcal{H} are then the orbits of this dynamical system. A distinguished feature of a Cauchy horizon when

viewed as a dynamical system is that the vector field T is nowhere vanishing, so none of the orbits of T are fixed points.

A number of issues arise in examining the behavior of the orbits of a given dynamical system (Σ^n, X) . Of primary interest here is whether or not (Σ^n, X) contains any periodic orbits (*i.e.*, orbits which pass repeatedly through the same point). We shall say that a periodic orbit λ is an *attractor* if all the nearby orbits approach it, and a *repeller* if they all move away. (In general, of course, a periodic orbit is neither a repeller nor an attractor.)

We wish now to briefly describe some specific examples of dynamical systems which we will find useful in our discussion of the dynamics of Cauchy horizons:

Example 1: Let Σ^2 be any two-dimensional manifold, and let ψ be any diffeomorphism from Σ^2 to itself. Let us recall the *suspension* construction [5] of a three-dimensional dynamical system which has global transverse section Σ^2 and has Poincaré map ψ : For the manifold Σ^3 of this dynamical system, one chooses the twisted product $\Sigma^2 \times_{\psi} S^1$, which is defined by quotienting $\Sigma^2 \times \mathbb{R}$ by the map

$$\begin{aligned}\Psi : \Sigma^2 \times \mathbb{R} &\rightarrow \Sigma^2 \times \mathbb{R}, \\ (p, s) &\mapsto (\psi^{-1}(p), s + 1).\end{aligned}$$

So $\Sigma^3 = \Sigma^2 \times_{\psi} S^1 \equiv \{\Sigma^2 \times \mathbb{R}\}/\Psi$. Then for the vector field X of the dynamical system, one chooses $X = \rho_*(\partial/\partial s)$, where ρ is the natural projection map $\rho : \Sigma^2 \times \mathbb{R} \rightarrow \Sigma^2 \times_{\psi} S^1$ associated with the definition of the twisted product $\Sigma^2 \times_{\psi} S^1$, and $\partial/\partial s$ is the vector field tangent to the \mathbb{R} factor of $\Sigma^2 \times \mathbb{R}$. One easily verifies that $(\Sigma^3, X) = (\Sigma^2 \times_{\psi} S^1, \rho_*(\frac{\partial}{\partial s}))$ has global transverse sections diffeomorphic to Σ^2 , and that ψ is the corresponding Poincaré map.

Now let L be any 2×2 matrix with integer entries, unit determinant, and eigenvalues with nonunit absolute value — *e.g.*, $L = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$. Based on the lattice quotient definition of the two-torus, any such matrix L defines a diffeomorphism $\psi_L : T^2 \rightarrow T^2$ of the two-torus to itself in a standard way. Then, using the suspension construction described above, we obtain for any such L a corresponding dynamical system (Σ_L^3, X_L) with ψ_L for its Poincaré map.

One verifies (*cf. e.g.* [6][pp.156–159]) that for any choice of L , the dynamical system (Σ_L^3, X_L) on the compact manifold Σ_L^3 (with nowhere vanishing generator X_L) has the following properties:

1. X_L has a countable infinity of periodic orbits. The set of all points $p \in \Sigma_L^3$ which lie on periodic orbits of X_L is dense in Σ_L^3 .
2. There are no attracting or repelling periodic orbits.
3. (Σ_L^3, X_L) is *ergodic* (*cf. e.g.* [7][Ex.5,p.19]).
4. (Σ_L^3, X_L) is *structurally stable*, so that all the properties here are preserved under all sufficiently small C^1 perturbations of the vector field X_L .

The dynamical systems (Σ_L^3, X_L) will be useful for building (stable) families of space-times which have compact Cauchy horizons with nonfountain-like generator dynamics (*cf. Section 3*).

Example 2: Consider⁴ a so-called DA diffeomorphism $\psi_{\text{DA}} : T^2 \rightarrow T^2$, as defined by Smale [5] (cf. also [6]). We do not wish to describe ψ_{DA} in detail; however, we wish to note the following. Let $(\Sigma_{\text{DA}}^3, X_{\text{DA}})$ be obtained by suspension of ψ_{DA} . Then the dynamical system $(\Sigma_{\text{DA}}^3, X_{\text{DA}})$ (with nowhere vanishing generator X_{DA}) exhibits the following properties [6][pp.165–169]:

1. There is one repelling orbit Γ ; there are no attracting orbits.
2. There is a *non-periodic* attracting set Λ (“strange attractor”), which is locally the product of \mathbb{R} with a Cantor set. Almost every orbit asymptotically approaches Λ , when followed to the future. Λ contains a countable infinity of periodic orbits (none of which are attractors or repellers).
3. The existence and properties of the attracting set Λ above are preserved under all sufficiently small smooth perturbations of the vector field X_{DA} .
4. There exists a neighborhood \mathcal{V} of the repelling orbit Γ such that an *arbitrary* perturbation of X_{DA} supported in \mathcal{V} will *not* affect the existence and the “chaotic” character of the attracting set Λ . (Such a perturbation might lead to a different basin of attraction of Λ . The new basin of attraction will nevertheless still have nonzero measure.)

We will use Example 2 (and some cutting and pasting) to build spacetimes containing compactly generated, noncompact, “asymptotically flat” Cauchy horizons with nonfountain-like generator dynamics (cf. Section 4).

3. COMPACT CAUCHY HORIZONS WITH NONFOUNTAIN-LIKE DYNAMICS

In this section we shall show that there exist smooth compact Cauchy horizons with no attracting periodic orbits. [Since a “fountain”, as defined in the Introduction, is precisely an attracting periodic orbit, the existence of spacetimes with Cauchy horizons with nonfountain-like behaviour immediately follows.] We have the following:

Proposition 3.1 *Let (Σ^3, X) be any dynamical system with Σ^3 compact and X nowhere vanishing. There exists a spacetime (M^4, g) (not necessarily satisfying any field equations and/or energy conditions) containing a Cauchy horizon \mathcal{H} which is diffeomorphic to Σ^3 , and such that the generators of \mathcal{H} are tangent to the orbits of X .*

We divide the proof into two main steps, the first of which involves proving the following Lemma:

Lemma 3.2 *Let (Σ^3, X) be any dynamical system with Σ^3 compact and X nowhere vanishing. Consider a spacetime (M^4, g) with $M^4 = \Sigma^3 \times (-\mu, \mu)$ for some $\mu > 0$, and let Z be a vector field on M^4 such that $Z|_{\Sigma^3 \times \{0\}} = X$. Suppose moreover that the following hold:*

1. $g(Z, Z)|_{\Sigma^3 \times \{0\}} = 0$,

⁴We are grateful to C. Robinson for pointing out this example to us.

2. $g^{-1}(dt, dt) < 0$ for all $t < 0$, where t parametrizes the interval $(-\mu, \mu)$.

Then:

1. $(\tilde{M}^4, \tilde{g}) \equiv (\Sigma^3 \times (-\mu, 0), g|_{\tilde{M}^4})$ is globally hyperbolic,
2. $\Sigma^3 \times \{0\}$ is a future Cauchy horizon for (\tilde{M}^4, \tilde{g}) in (M^4, g) , and
3. $Z|_{\Sigma^3 \times \{0\}} = X$ is tangent to the null generators of that Cauchy horizon.

Proof of Lemma: It follows from hypothesis 2 of this Lemma that the function

$$\begin{aligned} T : M^4 &= \Sigma^3 \times (-\mu, \mu) \rightarrow \mathbb{R} \\ (p, t) &\mapsto t \end{aligned}$$

is a time function on \tilde{M}^4 . Hence we know from Theorem 8.2.2 in Wald [8] that the spacetime (\tilde{M}^4, \tilde{g}) as defined above is stably causal, and further (cf. the Corollary on p. 199 of [8]) that it is strongly causal. Now to show that (\tilde{M}^4, \tilde{g}) is globally hyperbolic, it is sufficient (cf. p. 206 of Hawking and Ellis [4]) to verify that in addition one has $J^+(p) \cap J^-(q)$ compact for every $p, q \in \tilde{M}^4$, where $J^+(p)$ is the closure of the future of p in \tilde{M}^4 , and $J^-(q)$ is the closure of the past of q in \tilde{M}^4 . But $J^+(p) \cap J^-(q)$ is certainly closed and it is also the subset of a compact region $\Sigma^3 \times [T(p), T(q)] \subset \tilde{M}^4$. Hence $J^+(p) \cap J^-(q)$ is compact, and it follows that (\tilde{M}^4, \tilde{g}) is globally hyperbolic⁵.

Now $Z|_{\mathcal{H}}$ is nowhere vanishing, tangent to \mathcal{H} and null. It follows that the integral curves of $Z|_{\mathcal{H}}$ are causal curves which never leave \mathcal{H} . Hence no subset of M^4 containing \tilde{M}^4 and larger than \tilde{M}^4 can be globally hyperbolic, and consequently \mathcal{H} is a Cauchy horizon. By [4] there is a unique null direction tangent to each point of a smooth Cauchy horizon, with the null generators being tangent to this direction, so it must be that $Z|_{\mathcal{H}}$ is tangent to the null generators at each point of \mathcal{H} . \square

Proof of Proposition 3.1: By Lemma 3.2 all we need to do now is show that for any dynamical system (Σ^3, X) with Σ^3 compact and X nonvanishing, we can always find a spacetime (M^4, g) which satisfies the hypotheses of the Lemma. So let μ be any positive real number, let t parametrize the interval $(-\mu, \mu)$ and set $M^4 = \Sigma^3 \times (-\mu, \mu)$. The vector field X , defined in an obvious way on $\Sigma^3 \times \{0\}$, may now be Lie-dragged along the flow of $\partial/\partial t$ to define the vector field Z on M^4 . By construction we have $Z|_{\Sigma^3 \times \{0\}} = X$. Note that this construction also guarantees that $dt(Z) = 0$. To construct the appropriate spacetime metric, we first arbitrarily choose the following three fields:

1. Let $\phi : (-\mu, \mu) \rightarrow \mathbb{R}$ be any monotonically decreasing function such that $\phi(0) = 0$. From ϕ , we construct $\chi : M^4 \rightarrow \mathbb{R}$ by setting $\chi(p, t) = \phi(t)$.
2. Let β be any one-form on M^4 such that $\beta(Z) = 1$ and $\beta(\partial/\partial t) = 0$.
3. Let γ be any Riemannian metric on Σ^3 ; for $V, W \in T\Sigma$ set

$$\begin{aligned} \tilde{\gamma}(V, W) &\equiv \gamma(V, W) - \frac{1}{2}\{\gamma(X, V)\beta(W) + \gamma(X, W)\beta(V)\}, \\ \tilde{\gamma}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) &\equiv \tilde{\gamma}\left(\frac{\partial}{\partial t}, W\right) \equiv 0 \end{aligned}$$

⁵Note that this argument shows, that a spatially compact stably causal spacetime is necessarily globally hyperbolic.

(note that $\tilde{\gamma}(X, \cdot) = 0$); and finally define ν as a symmetric $\binom{0}{2}$ tensor field on M^4 by dragging $\tilde{\gamma}$ along $\frac{\partial}{\partial t}$. [Here β has been identified with a form on Σ in the obvious way.] Note that it follows from this definition that $\nu(Z, Z) = 0$ and $\nu(\partial/\partial t, \partial/\partial t) = 0$.

Using χ , β , and ν , we define

$$g = \chi\beta \otimes \beta + dt \otimes \beta + \beta \otimes dt + \nu. \quad (1)$$

We verify immediately from the properties of χ , β , and ν and from the definition of Z that $g(Z, Z) = \chi$ everywhere on M , and in particular $g(Z, Z)|_{\Sigma^3 \times \{0\}} = 0$, so hypothesis 1 of Lemma 3.2 is satisfied. To verify hypothesis 2 it is useful to set up local coordinates (x, y, z, t) such that $\beta = dz$, $Z = \partial/\partial z$; it follows that $\nu = \nu_{xx}dx^2 + 2\nu_{xy}dxdy + \nu_{yy}dy$. Then the components of the metric g take the matrix form

$$g_{\alpha\beta} = \begin{pmatrix} \nu_{xx} & \nu_{xy} & 0 & 0 \\ \nu_{yx} & \nu_{yy} & 0 & 0 \\ 0 & 0 & \chi & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

From this matrix representation, we see that g is indeed a Lorentz metric (nondegenerate, signature $++-$) and we calculate the matrix representation of the inverse metric:

$$(g^{-1})^{\alpha\beta} = \begin{pmatrix} \frac{1}{\det \nu} \nu_{yy} & \frac{-1}{\det \nu} \nu_{xy} & 0 & 0 \\ \frac{-1}{\det \nu} \nu_{yx} & \frac{1}{\det \nu} \nu_{xx} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -\chi \end{pmatrix}.$$

We see that $g^{-1}(dt, dt) = -\chi$, which implies that $g^{-1}(dt, dt) < 0$ for $t < 0$, as required by hypothesis 2 of Lemma 3.2. So the spacetime (M^4, g) which we have constructed (from the dynamical system (Σ^3, X)) satisfies all of the hypotheses of the Lemma, thus completing the proof of the Proposition. \square

Using the example dynamical systems from Section 2, together with this Proposition (and some of the constructions outlined in its proof), we can easily construct a large numbers of spacetimes containing compact Cauchy horizons with nonfountain-like dynamics. Here the only essential restriction is, that the vector field X generating the dynamical system be nowhere vanishing — this excludes examples like e.g. (a compactified version of) the Lorenz attractor [9] or of the geometric model thereof [10], but clearly allows for interesting dynamics. Models with e.g. “horseshoes” can be constructed on $S^2 \times S^1$ using the periodically perturbed nonlinear pendulum equation or the periodically perturbed Duffing equation.

We wish to stress that these examples can be constructed in such a way that the nontrivial properties of the dynamics are *stable under small smooth variations of the metric*. For example, let (Σ^3, X) be the Anosov flow discussed in Example 1 of the previous Section. The metric g constructed in the proof of Proposition 3.1 can be chosen to satisfy the stability criterion of [11], so that small smooth variations of the metric will lead to small C^k variations⁶ of the Cauchy horizon. This in turn will lead to a small C^{k-1} variation of the field of null tangents to the generators, and the stability of the resulting dynamical systems follows from stability of Anosov flows.

⁶ k here may be made arbitrarily large (but probably *not* $k = \infty$) by appropriately choosing g .

Note that all examples discussed so far have Σ^3 defined as a twisted product of a two-dimensional manifold with the circle. Do all compact Cauchy horizons with nonfountain-like behavior have this sort of topology? Certainly not. In the next Section we shall see how to construct Cauchy horizons with interesting dynamical behaviour of the generators by using the connected sum operation. It would be of some interest to find out whether or not there are spacetimes with nonfountain-like behavior in a Cauchy horizon of arbitrary (compact, three-dimensional) topology.

4. NONCOMPACT COMPACTLY GENERATED HORIZONS

It is relatively easy to construct a spacetime (M^4, g) which has a compactly-generated but noncompact Cauchy horizon with nonfountain-like dynamics. First, one chooses a compact dynamical system (Σ^3, X) which has nonfountain-like dynamics and also has a repelling periodic orbit: the DA system as discussed in example 2 will do. Then, one uses Proposition 3.1 to construct a spacetime containing a Cauchy horizon diffeomorphic to Σ^3 with generators matching the orbits of X . Finally one removes the repelling orbit from the horizon in the spacetime. The Cauchy horizon \mathcal{H} of the resulting spacetime is clearly not compact. On the other hand, one verifies that \mathcal{H} is compactly-generated by noting that if one defines the set $\mathcal{K} = \mathcal{H} \setminus \tilde{S}$, where \tilde{S} is a small open thickening of the removed orbit, then since the removed orbit was repelling, all past-directed null generators of \mathcal{H} enter and remain in \mathcal{K} , which is compact. [Note that this example shows that the inequality $f \geq 0$, which according to [1] holds for any periodic generator of a Cauchy horizon, is not correct.]

The above example is rather artificial, and it is natural to enquire about the existence of *smooth* compactly generated horizons in asymptotically flat spacetimes. By way of example, consider $M = \mathbb{R}^4$ with a metric g which is the standard Minkowski metric outside of a compact set \mathcal{C} . Let us moreover assume that there exist periodic time-like curves in \mathcal{C} . [An explicit example of such a spacetime can be found in [1].] M will have a Cauchy horizon \mathcal{H}^+ , which is the boundary of the domain of dependence of any standard $t = \text{const}$ plane lying in $M \setminus J^+(\mathcal{C})$. Now \mathcal{H}^+ can be “sandwiched” between $\partial J^+(p)$ and $\partial J^+(q)$, where p, q are any two points such that $\mathcal{C} \subset J^+(p)$, and $\mathcal{C} \cap J^+(q) = \emptyset$; by “sandwiched” here we mean that $\mathcal{H}^+ \subset \{I^-(\partial J^+(q)) \cap I^+(\partial J^+(p))\}$. It is then easily seen that for all $R \in \mathbb{R}$ large enough the world tube $\mathcal{T} = \{(t, \vec{x}) : t \in \mathbb{R}, |\vec{x}| = R\}$ intersects each of the generators of \mathcal{H}^+ transversally.

Based on this example, we shall say that a compactly generated Cauchy horizon \mathcal{H} in an orientable and time-orientable spacetime (M, g) is of *asymptotically flat type* if the boundary set $\partial(\mathcal{K} \cap \mathcal{H})$ consists of a finite number I of spheres S_i , with each generator of $\mathcal{H} \setminus \mathcal{K}$ intersecting one of the S_i 's transversally. Here \mathcal{K} is one of the compact sets in M which characterizes \mathcal{H} as compactly generated. It is easy to convince oneself that the behaviour described in this definition should occur, *e.g.*, for compactly generated Cauchy horizons in spacetimes which admit a sufficiently regular compactification in lightlike directions (the number I above corresponds then to the number of connected components of Scri).

We would like to find spacetimes with compactly-generated, asymptotically flat type Cauchy horizons with nonfountain-like dynamics. The following result is an obstacle to the construction of such spacetimes:

Proposition 4.1 *Let \mathcal{H}^+ be a compactly generated future Cauchy horizon of asymptotically flat type. Then the field X of directions tangent to the generators of \mathcal{H}^+ cannot be continuous.*

Proof: Suppose that the field X is continuous. Now consider the compact manifold $\hat{\mathcal{H}}$ constructed by adding a point p_∞^i to each of the “asymptotic ends” $S_i \times \mathbb{R} \subset \mathcal{H}^+$. We can deform the field of generators on each of the ends $S_i \times \mathbb{R}$ to obtain a continuous vector field \hat{X} on $\hat{\mathcal{H}}$ which is nowhere vanishing except at the points p_∞^i . At each of those points the index of \hat{X} will be equal to +1; consequently the index of \hat{X} will be equal to $I \neq 0$. Note that $\hat{\mathcal{H}}$ is orientable because (M, g) has been assumed to be time-orientable and orientable. This, however, contradicts the fact that the index of a continuous vector field on a compact, three-dimensional, orientable manifold vanishes. \square

This result makes it difficult to systematically study the dynamics of the null generators in spacetimes containing compactly generated Cauchy horizons of asymptotically flat type. In particular, the construction carried out in Proposition 3.1 encounters various obstacles. However, as it has been suggested [2, 12, 1] that there exist compactly generated Cauchy horizons \mathcal{H} of asymptotically flat type which are smooth on an open dense set \mathcal{U} (the complement of which has zero measure), we believe the following result should be of interest:⁷

Proposition 4.2 *Let $\mathcal{H}^+ = \partial^+ \mathcal{D}(\Sigma)$ be a future Cauchy horizon in a space-time (M, g) . Suppose that there exists an open subset \mathcal{U} of \mathcal{H}^+ such that \mathcal{U} is a smooth submanifold of M . Suppose moreover that there exists a smooth time function τ on $\mathcal{D}(\Sigma)$ such that $\lim_{\mathcal{D}(\Sigma) \ni p \rightarrow \mathcal{U}} \nabla \tau(p)$ exists, and is a smooth, nowhere vanishing vector field on \mathcal{U} . Then there exists a space-time (M', g') with a future Cauchy horizon \mathcal{H}' diffeomorphic to $\mathcal{H}^+ \# \Sigma_{\text{DA}}^3$ (where Σ_{DA}^3 is the manifold discussed in Example 2, Section 2), and with non-trivial long-time dynamics of the generators of \mathcal{H}' . [Here $\#$ denotes the connected sum.] Moreover, \mathcal{H}' will share certain overall properties of \mathcal{H}^+ ; in particular if \mathcal{H}^+ is compact, or compactly generated, or of asymptotically flat-type, then the same will be true of \mathcal{H}' .*

Remarks: We believe that the inclusion in Proposition 4.2 of the hypothesis that the function τ exists should be unnecessary, for the following reasons:

1. We consider it likely that the remaining conditions of Proposition 4.2 are sufficient to guarantee that such a function can be constructed.
2. We have written the proof below in such a way that the existence of the function τ is essentially used in one place only. We believe that it should be possible to replace that step of the argument by one which does *not* require the existence of the function τ .

Proof: Let Σ be a partial Cauchy surface in (M, g) such that $\mathcal{H}^+ = \partial^+ \mathcal{D}(\Sigma)$. Replacing M by a subset thereof if necessary we may assume that $\mathcal{H}^- \equiv \partial^- \mathcal{D}(\Sigma) = \emptyset$. Let \mathcal{U} be a smooth subset of \mathcal{H}^+ . Passing to a subset of \mathcal{U} if necessary we may without loss of generality assume that: 1) the closure $\bar{\mathcal{U}}$ of \mathcal{U} is compact, and 2) that the generators of

⁷Here we define $\partial^+ \mathcal{D}(\Sigma) = \overline{\mathcal{D}^+(\Sigma)} \setminus (\mathcal{D}^+(\Sigma) \cup \Sigma)$, similarly for $\partial^- \mathcal{D}(\Sigma)$. We use the convention in which the domains of dependence are *open* sets; in particular they do *not* include the Cauchy horizons.

\mathcal{H}^+ have a cross-section S in \mathcal{U} , and 3) that $\mathcal{U} \approx S \times (-1, 1)$, with S being a smooth two-dimensional embedded submanifold. We claim that we can find a defining function φ for \mathcal{U} , defined on a conditionally compact neighborhood \mathcal{O} of \mathcal{U} , such that $\varphi|_{\mathcal{O} \cap \mathcal{D}^+(\Sigma)}$ is a time function. [Recall that $\varphi : \mathcal{O} \rightarrow \mathbb{R}$ is a defining function for \mathcal{U} if $d\varphi$ is nowhere vanishing on \mathcal{U} and if we have $p \in \mathcal{U} \cap \mathcal{O} \Leftrightarrow \varphi(p) = 0$.] If we have a time function τ , as assumed in the hypotheses of Proposition 4.2, we set $\varphi = \tau|_{\mathcal{O} \cap \mathcal{D}^+(\Sigma)}$, and we are done.

[Had we not made the assumption of the existence of τ , the existence of φ could be established as follows: Let ψ be any defining function for \mathcal{U} defined on some neighborhood \mathcal{O} , and let X be any future-directed timelike vector field on \mathcal{O} . If

$$[X^\mu \nabla_\mu (\nabla^\nu \psi \nabla_\nu \psi)]|_{\mathcal{U}} > 0, \quad (2)$$

then passing to a subset of \mathcal{O} if necessary we shall have $\nabla^\nu \psi \nabla_\nu \psi|_{\mathcal{D}^+(\Sigma) \cap \mathcal{O}} < 0$, and then setting $\varphi = \psi$ we are done. If (2) does not hold, consider any smooth function α on \mathcal{O} ; we have

$$X^\mu \nabla_\mu (\nabla^\nu (\alpha \psi) \nabla_\nu (\alpha \psi))|_{\mathcal{U}} = \alpha^2 [X^\mu \nabla_\mu (\nabla^\nu \psi \nabla_\nu \psi) + X^\nu \nabla_\nu \psi \nabla^\mu \psi \nabla_\mu (\log^2 \alpha)]. \quad (3)$$

Note that $X^\nu \nabla_\nu \psi$ is nowhere vanishing on \mathcal{U} , as X^ν is time-like and $\nabla^\nu \psi$ is null. Let $\hat{\alpha} : \mathcal{U} \rightarrow \mathbb{R}$ be any strictly positive solution of the equation

$$[\nabla^\mu \psi \nabla_\mu (\log^2 \hat{\alpha})] = (X^\mu \nabla_\mu \psi)^{-1} [1 - X^\mu \nabla_\mu (\nabla^\nu \psi \nabla_\nu \psi)]|_{\mathcal{U}},$$

and let α be any strictly positive extension of $\hat{\alpha}$ to \mathcal{O} . Setting $\varphi = \alpha \psi$ the desired defining function then follows. Passing to a subset of \mathcal{O} we may moreover assume that $d\varphi$ is nowhere vanishing on \mathcal{O} . Changing φ to $-\varphi$ if necessary we may suppose that $\nabla^\nu \varphi$ is past-directed on $\mathcal{D}^+(\Sigma) \cap \mathcal{O}$.]

Let $\mathcal{B}_{4\rho} \subset \mathcal{U}$ be a closed coordinate ball of radius 4ρ covered by coordinates x^i , with the x^i 's chosen so that $g^{\mu\nu} \varphi_{,\mu}|_{\mathcal{H}} = \frac{\partial}{\partial x^3}$, and with $\mathcal{B}_{4\rho}$ compact in \mathcal{U} . If we choose a timelike future directed vector field T on \mathcal{O} , then by dragging the coordinates x^i along the integral curves of T we obtain a coordinate system $(x^0, x^i) = (\varphi, x^i)$ on a compact set $[-\delta, \delta] \times \mathcal{B}_{4\rho}$, for some $\delta > 0$. Since φ is a time function on $[-\delta, 0] \times \mathcal{B}_{4\rho}$, the sets $\{s\} \times \mathcal{B}_{4\rho}$ are spacelike for $s \in [-\delta, 0)$. In this coordinate system the metric takes the form

$$g_{\mu\nu} dx^\mu dx^\nu = 2g_{30} dx^0 dx^3 + g_{00}(dx^0)^2 + 2g_{0A} dx^0 dx^A + g_{AB} dx^A dx^B + O(\varphi), \quad (4)$$

where the labels A, B run over 1, 2, and where $O(\varphi)$ indicates terms which vanish at least as fast as $|\varphi|$ for small values of $|\varphi|$. Consequently,

$$\det g = -(g_{30})^2 \det(g_{AB}) + O(\varphi), \quad (5)$$

from which it follows that if δ is sufficiently small, then g_{30} does not change sign. From the above construction, it follows that in fact g_{30} is positive.

Let t be any strictly negative time function on $\mathcal{D}(\Sigma)$ such that 1) the level sets of t are Cauchy surfaces for $\mathcal{D}(\Sigma)$, and 2) $t(p) \rightarrow 0$ as $p \rightarrow \partial^+ \mathcal{D}(\Sigma)$. Then set

$$\epsilon = \inf -t(p) \mid p \in \{-\delta\} \times \mathcal{B}_{2\rho}.$$

Now, consider the spacetime region (M_1, g^1) with

$$M_1 = M \setminus (J^-(-\delta) \times \mathcal{B}_{2\rho}) \cup (\{\delta\} \times \mathcal{B}_{2\rho}), \quad g^1 = g|_{M_1}. \quad (6)$$

Clearly, $\Sigma = \{t = -\epsilon/2\}$ is a partial Cauchy surface in M_1 such that $\partial^+ \mathcal{D}(\Sigma; M_1) = \mathcal{H}^+$ is a future Cauchy horizon for Σ . Here we use the notation $\mathcal{D}(\Sigma; M_1)$ for the domain of dependence of Σ in (M_1, g^1) ; we shall use a similar convention for J^\pm , etc.

Let Σ_{DA}^3 be the manifold discussed in Example 2, Section 2. Let Γ be the repelling orbit and \mathcal{V} be the designated neighborhood of Γ as discussed in that Example, and let $\mathcal{B}_{4\rho}^1 \subset \mathcal{V}$ be a closed coordinate ball covered by coordinates y^i . Finally let ψ be the inversion map:

$$\mathcal{U} \supset \hat{\mathcal{B}}_{2\rho} \setminus \mathcal{B}_{\rho/2} \ni x^i \xrightarrow{\psi} y^i = -\frac{x^i}{r(x)^2} \in \hat{\mathcal{B}}_{2\rho}^1 \setminus \mathcal{B}_{\rho/2}^1 \subset \Sigma_{\text{DA}}^3 ,$$

where $r(x) = \sqrt{\sum(x^i)^2}$. It is easily shown that one can find a nowhere vanishing vector field X on $\Sigma_{\text{DA}}^3 \setminus \mathcal{B}_{\rho}^1$ such that

$$(\psi^{-1})_* X \Big|_{\hat{\mathcal{B}}_{2\rho} \setminus \hat{\mathcal{B}}_\rho} = \nabla \varphi ,$$

and

$$X \Big|_{\Sigma_{\text{DA}}^3 \setminus \mathcal{B}_{\rho/2}^1} = X_{\text{DA}} ,$$

where X_{DA} is the generator of the DA flow discussed in Section 2, and where we have used “hats” to denote the interior of a set: $\hat{\mathcal{B}}_\rho = \text{int } \mathcal{B}_\rho$, etc. Now let g^{DA} be any Lorentzian metric constructed on $M_{\text{DA}} \equiv (-\delta, \delta) \times \{\Sigma_{\text{DA}}^3 \setminus \mathcal{B}_{\rho/2}^1\}$ as described in the Proof of Proposition 3.1. On $(-\delta, 0) \times \{\Sigma_{\text{DA}}^3 \setminus \mathcal{B}_{\rho/2}^1\} \subset M_{\text{DA}}$, we can define a time function y^0 by

$$y^0(s, p) = s .$$

Based on the map Ψ we define

$$(-\delta, \delta) \times \{\hat{\mathcal{B}}_{2\rho} \setminus \mathcal{B}_{\rho/2}\} \ni (t, p) \rightarrow \Psi(t, p) = (t, \psi(p)) \in (-\delta, \delta) \times \{\hat{\mathcal{B}}_{2\rho}^1 \setminus \mathcal{B}_{\rho/2}^1\} .$$

Then, in local coordinates on $(-\delta, \delta) \times \{\hat{\mathcal{B}}_{2\rho} \setminus \mathcal{B}_{\rho/2}\}$ the metric $\Psi^* g^{\text{DA}}$ takes the form

$$g_{\mu\nu}^{\text{DA}} dx^\mu dx^\nu = 2\beta_\mu dx^0 dx^\mu + g_{ij}^{\text{DA}} dx^i dx^j , \quad (7)$$

where

$$\beta_0 = 0, \quad \beta_i dx^i = \psi^* \beta .$$

On $\hat{\mathcal{B}}_{2\rho} \setminus \mathcal{B}_\rho$ we have

$$\beta_3 = \beta_\mu \nabla^\mu \varphi = \langle \Psi^* \beta, \nabla \varphi \rangle = \langle \beta, \Psi_* \nabla \varphi \rangle = \langle \beta, X \rangle = 1 . \quad (8)$$

Since the metric g^{DA} is y^0 -independent, (8) actually holds on $(-\delta, \delta) \times \{\hat{\mathcal{B}}_{2\rho} \setminus \mathcal{B}_\rho\}$. A similar calculation shows that

$$g_{3i}^{\text{DA}} = 0 \quad (9)$$

on $(-\delta, \delta) \times \{\hat{\mathcal{B}}_{2\rho} \setminus \mathcal{B}_\rho\}$. Now let $\phi \in C^\infty(\mathbb{R}^4)$ be any non-negative function such that $\phi = 0$ in $\mathbb{R} \times \mathcal{B}_\rho$, and $\phi = 1$ in $\mathbb{R}^4 \setminus (\mathbb{R} \times \mathcal{B}_{2\rho})$. On M_1 we may define the smooth metric g^2 to coincide with g^1 on $M \setminus \{(-\delta, \delta) \times \mathcal{B}_{2\rho}\}$, and to be given by

$$g_{\mu\nu}^2 dx^\mu dx^\nu = \phi g_{\mu\nu}^{\text{DA}} dx^\mu dx^\nu + (1 - \phi) g_{\mu\nu}^{\text{DA}} dx^\mu dx^\nu \quad (10)$$

on $(-\delta, \delta) \times \{\mathcal{B}_{2\rho} \setminus \mathcal{B}_{\rho/2}\}$ (it is easily seen from eqs. (4), (7)–(10) and from eq. (5) with g^2 substituted for g , that (10) indeed defines a Lorentzian metric). Note that $\varphi = x^0$ is still a time function for this new metric in $(-\delta, 0) \times \{\mathcal{B}_{2\rho} \setminus \mathcal{B}_{\rho/2}\}$.

The desired spacetime M' will now be obtained by gluing together (M_1, g^2) and $(M_{\text{DA}}, g^{\text{DA}})$: Specifically, we choose

$$M' = \left[\{M_1 \setminus [(-\delta, \delta) \times \hat{\mathcal{B}}_{\rho/2}]\} \sqcup M_{\text{DA}} \right] / \Psi . \quad (11)$$

Since the metrics g^2 and $\Psi^* g^{\text{DA}}$ coincide on $\mathcal{B}_\rho \setminus \mathcal{B}_{\rho/2}$, a metric g' can be defined on M' in the obvious way. There is a natural identification between points in $M_1 \setminus [(-\delta, \delta) \times \hat{\mathcal{B}}_{\rho/2}]$ and an appropriate subset of M' , and similarly for points in M_{DA} , with another subset of M' . We shall use this identification without mentioning it explicitly in what follows. Let us note that the function φ' , defined as

$$\varphi'(p) = \begin{cases} \varphi(p) (= x^0(p)), & p \in (-\delta, 0] \times [\mathcal{B}_{4\rho} \setminus \hat{\mathcal{B}}_{\rho/2}], \\ y^0(p), & p \in (-\delta, 0] \times [\Sigma_{\text{DA}}^3 \setminus \hat{\mathcal{B}}_{\rho/2}^1], \end{cases}$$

is a smooth time function on the interior of the set on which it has been defined.

We now claim that the submanifold \tilde{M} of M' defined by

$$\tilde{M} = \left[\{\mathcal{D}(\Sigma; M_1) \setminus [(-\delta, 0) \times \hat{\mathcal{B}}_{\rho/2}]\} \sqcup \{(-\delta, 0) \times [\Sigma_{\text{DA}}^3 \setminus \hat{\mathcal{B}}_{\rho/2}^1]\} \right] / \Psi ,$$

with the metric obtained from g' by restriction, is globally hyperbolic. First, we wish to show that for all $p, q \in \tilde{M}$, the set $J^+(p; \tilde{M}) \cap J^-(q; \tilde{M})$ is compact. To do this it is convenient to consider various cases, according to whether $p \in \mathcal{D}(\Sigma; M_1) \setminus [(-\delta, 0) \times \hat{\mathcal{B}}_\rho]$, $p \in (-\delta, 0) \times [\hat{\mathcal{B}}_\rho \setminus \mathcal{B}_{\rho/2}]$, or $p \in (-\delta, 0) \times [\Sigma_{\text{DA}}^3 \setminus \hat{\mathcal{B}}_{\rho/2}^1]$, similarly for q . Suppose, e.g., that $p, q \in M_1 \setminus [(-\delta, 0) \times \hat{\mathcal{B}}_\rho]$. We define

$$K = J^+(p; M_1) \cap J^-(q; M_1) \cap [(-\delta, 0) \times \partial \mathcal{B}_\rho] .$$

K is easily seen to be compact by global hyperbolicity of $(\mathcal{D}(\Sigma; M_1), g^1)$. If $K = \emptyset$ we have $J^+(p; \tilde{M}) \cap J^-(q; \tilde{M}) = J^+(p; M_1) \cap J^-(q; M_1)$ and we are done; otherwise we have

$$-\delta < s_- = \inf \varphi(p) < s_+ = \sup \varphi(p) < 0$$

where the sup and the inf are taken over $p \in K$. Since we have a time function φ' on $\tilde{M} \setminus \{M_1 \setminus [(-\delta, 0) \times \hat{\mathcal{B}}_\rho]\}$ which agrees with φ on K , it follows that $J^+(p; \tilde{M}) \cap J^-(q; \tilde{M})$ can be covered by the compact sets $J^+(p; M_1) \cap J^-(q; M_1)$, $[-s_-, s_+] \times [\mathcal{B}_\rho \setminus \hat{\mathcal{B}}_{\rho/2}]$ and $[-s_-, s_+] \times [\Sigma_{\text{DA}}^3 \setminus \hat{\mathcal{B}}_{\rho/2}^1]$. Using similar arguments one shows compactness of $J^+(p; \tilde{M}) \cap J^-(q; \tilde{M})$ for the remaining cases.

To prove strong causality of \tilde{M} , we use the existence of the time function τ on $\mathcal{D}(\Sigma, M)$: Indeed, the function τ' defined by:

$$\tau'(p) = \begin{cases} \tau(p), & p \in M_1 \setminus [(-\delta, \delta) \times \hat{\mathcal{B}}_{\rho/2}], \\ y^0(p), & p \in M_{\text{DA}}, \end{cases}$$

is a smooth time function on \tilde{M} . This ensures strong causality of \tilde{M} , and global hyperbolicity of \tilde{M} follows. [Let us emphasize, that this is the only point at which the hypothesis of existence of τ is needed⁸ in our argument.]

We wish to show now that the set \mathcal{H}' , defined as the boundary of \tilde{M} in M' , is a Cauchy horizon. Note first that the generators of \mathcal{H}^+ in $\mathcal{B}_\rho \setminus \hat{\mathcal{B}}_{\rho/2}$ are the integral curves

⁸We believe that \tilde{M} as constructed here is strongly causal even without the assumption of existence of the function τ ; we have, however, not been able to prove this assertion.

of $\nabla\varphi = \nabla'\varphi$, where φ is the defining function for \mathcal{H}^+ on \mathcal{U} defined at the beginning of this proof, and ∇' is the gradient with respect to the metric g' . These curves are smoothly continued by the null (with respect to the metric g') integral curves of $\nabla'\varphi'$. Consider now any subset \check{M} of M' which contains \check{M} as a proper subset. It follows that $\check{M} \cap \mathcal{H}' \neq \emptyset$. Let $p \in \check{M} \cap \mathcal{H}'$. We claim that there exists a past directed causal curve λ through p which never enters \check{M} : If $p \in M_1$, then consider a generator Γ of \mathcal{H}^+ through p . If Γ never enters \mathcal{B}_ρ when followed backwards with respect to the time orientation, then the connected component of $\Gamma \cap \check{M}$ which contains p provides the desired curve λ . If Γ enters \mathcal{B}_ρ , let $\tilde{\Gamma}$ be the segment of Γ up to the point \tilde{p} where it first enters \mathcal{B}_ρ . $\tilde{\Gamma}$ can be smoothly continued at \tilde{p} by the integral curve Γ' of $\nabla'\varphi'$. If this curve never exits \mathcal{B}_ρ through the sphere $\partial\mathcal{B}_\rho$, we can set λ to be the connected component of $(\Gamma \cup \Gamma') \cap \check{M}$ which contains p . If it exits \mathcal{B}_ρ through the sphere $\partial\mathcal{B}_\rho$, then it can be smoothly continued by a segment Γ'' of a generator of \mathcal{H}^+ . If Γ'' never reenters \mathcal{B}_ρ , we define λ to be the connected component of $(\Gamma \cup \Gamma' \cup \Gamma'') \cap \check{M}$ that contains p . If Γ'' reenters \mathcal{B}_ρ when followed backwards with respect to the time orientation, we continue the procedure above to eventually obtain an inextendible curve λ through p . This shows the existence of an inextendible, past-directed causal curve $\lambda \subset \mathcal{H}' \cap \check{M}$ through all $p \in \mathcal{H}' \cap \check{M}$. Hence it follows that

$$\mathcal{H}' = \partial^+ \mathcal{D}(\Sigma'; M')$$

for some $\Sigma' \subset M'$.

Clearly we have

$$\mathcal{H}' \approx \mathcal{H}^+ \# \Sigma_{\text{DA}}^3,$$

where $\#$ denotes the connected sum. Moreover the generators of \mathcal{H}' coincide with

1. those of \mathcal{H}^+ on $\mathcal{H}^+ \setminus \mathcal{B}_\rho$, and with
2. the integral curves of the suspension of the DA-diffeomorphism on $\Sigma_{\text{DA}}^3 \setminus \mathcal{B}_{3\rho/2}^1$.

From what has been said in Example 2, Section 2, it follows that a “non-zero measure” set of generators of \mathcal{H}' will be attracted to a “strange attractor”, when followed backwards in time. \square

We expect that some of the examples constructed as in Proposition 4.2 are stable in the dynamical sense. However, we have no proof of this assertion. To establish stability one would need to prove that small smooth variations of the metric lead to small C^2 variations of the horizon on (perhaps an open subset of) \mathcal{U} . Now it is not difficult to show that, for an appropriately chosen (M, g) , the metric g' on M' can be constructed so that small variations of g' will indeed lead to small C^0 variations of the horizon \mathcal{H}' (*cf. e.g.* [13] for various results of this kind). The transition from C^0 to C^2 seems, however, to be a non-trivial matter. In particular, we have not been able to generalize the methods of [11] from compact Cauchy horizons with global cross-sections to noncompact Cauchy horizons.

ACKNOWLEDGEMENTS

We acknowledge a fruitful collaboration with J. Friedman at an early stage of the work on this paper. Useful discussions with B. Birnir, V. Moncrief, C. Robinson and K.

Thorne are also acknowledged. The authors are grateful to the Institute for Theoretical Physics of the University of California at Santa Barbara for hospitality and financial support (under NSF grant PHY89-04035) during part of the work on this paper. They were also supported in part by NSF grant PHY90-12301 to University of Oregon. Moreover, J.I. was supported in part by NSF grant PHY93-8117, while P.T.C. was supported in part by KBN grant # 2 1047 9101, by NATO and by the Alexander von Humboldt Foundation.

REFERENCES

- [1] Hawking, S.W., 1992, "Chronology protection conjecture", *Phys. Rev.*, **D46**, 603-611.
- [2] Thorne, K.S., 1993, "Closed time-like curves", in *General Relativity and Gravitation 13*, R.J. Gleiser, C.N. Kozameh, O.M. Moreschi, eds., (Bristol: Institute of Physics).
- [3] Visser, M., 1993, "Van Vleck determinants: traversable wormhole spacetimes", preprint, gr-qc/9311026.
- [4] Hawking, S.W., Ellis, G.F.R., 1973, *The Large Scale Structure of Space-time*, (Cambridge: Cambridge University Press).
- [5] Smale, S., 1967, "Differential dynamical systems", *Bull. AMS*, **73**, 747-817.
- [6] Palis, J., Jr., de Melo, W., 1982, *Geometric Dynamical Systems*, (New York: Springer).
- [7] Sinai, Ya.G., 1989, *Dynamical Systems*, Vol. II, (New York: Springer).
- [8] Wald, R.M., 1984, *General Relativity*, (Chicago: University of Chicago Press).
- [9] Sparrow, C., 1982, *The Lorenz equations: bifurcations, chaos and strange attractors*, Springer Lecture Notes in Math., Vol. 41.
- [10] Guckenheimer, J., Williams, R.F., 1979, "Structural stability of Lorenz attractors", *Publ. Math. IHES*, **50**, 59-72 .
- [11] Chruściel, P.T., Isenberg, J., 1993, "On stability of differentiability of Cauchy horizons", in preparation.
- [12] Friedman, J., Morris, M.S., Novikov, I., Echeverria, F., Glinkhamer, R., Thorne, K.S., Yurtsever, U., 1990, "Cauchy problem in spacetimes with closed timelike curves", *Phys. Rev.*, **D 42**, 1915-1930.
- [13] Beem, J., Królak, A., 1993, "Causality and Cauchy horizons", preprint.

COMPACT RELATIVISTIC SYSTEMS

CHAOS IN THE CASE OF TWO FIXED BLACK HOLES

G. Contopoulos

Dept. of Astronomy, University of Florida
Gainesville, FL 32611, USA
and Dept. of Astronomy, University of Athens
Panepistimiopolis, GR 15783 Athens, Greece

Abstract. We study the orbits of photons (null geodesics) and particles (time-like geodesics) in the field of two fixed black holes. The orbits of photons terminate either at the black holes M_1 and M_2 (types (I) and (II)), or at infinity (type (III)). The limits of these three types of orbits are Cantor sets, defined by the unstable periodic orbits, that form a set of measure zero. In the case of particles with elliptic energy there are some stable periodic orbits that trap around them both quasiperiodic orbits (type (IV)) and chaotic orbits that do not reach the black holes (type (V)). The stable periodic orbits may become unstable, producing an infinity of period doubling bifurcations. The main periodic orbits are two almost circular satellite orbits around each black hole, M_1 and M_2 , and almost elliptical orbits surrounding both black holes. Orbits crossing one of the satellite orbits inwards fall into the corresponding black hole. For certain values of the parameters one or both satellite orbits may not exist. Then the separation of the type (I) and/or type (II) orbits is done by some almost hyperbolic orbits. In the Newtonian case the satellite orbits around M_1 and M_2 never exist (except if $M_1 = 0$, or $M_2 = 0$ exactly). However we have many families of higher order periodic orbits that exist both in the relativistic and the Newtonian case. The differences between the corresponding orbits are large close to the black holes.

1. INTRODUCTION

The scattering of radiation from two fixed black holes was studied by Chandrasekhar (1989). The black holes are assumed to be of extreme Reissner-Nordström type, i.e.

their charges are such that their electrostatic repulsion is equal to their gravitational attraction. Otherwise the masses of the two black holes are arbitrary. Chandrasekhar found the perturbations of the metric of the two black holes, representing gravitational and electromagnetic radiation, that is either absorbed by the two black holes, or scattered to infinity. In an Appendix he gave the null geodesics of some photons on the meridian plane of the two black holes.

A systematic study of the geodesics of photons (null geodesics) and of particles (time-like geodesics) was made by Contopoulos (1990a, 1991). A comparison between the Newtonian and relativistic periodic orbits was given by Contopoulos and Papadaki (1992). Here we review the main results of these studies, including some more recent calculations.

The Lagrangian of two fixed black holes is (Chandrasekhar 1989)

$$2L = \dot{t}^2/U^2 - U^2 \left[(\sinh^2 \psi + \sin^2 \theta) (\dot{\psi}^2 + \dot{\theta}^2) + \sinh^2 \psi \sin^2 \theta \dot{\phi}^2 \right], \quad (1)$$

where

$$U = 1 + \frac{M_1}{r_1} + \frac{M_2}{r_2}, \quad (2)$$

$$r_1^2 = x^2 + y^2 + (z - 1)^2, \quad r_2^2 = x^2 + y^2 + (z + 1)^2, \quad (3)$$

$$x = \sinh \psi \sin \theta \cos \varphi, \quad y = \sinh \psi \sin \theta \sin \varphi, \quad z = \cosh \psi \cos \theta, \quad (4)$$

Here ψ, θ, ϕ are prolate spheroidal coordinates and the dots represent derivatives with respect to an affine parameter. This is the well known Majumdar (1947) - Papapetrou (1947) solution of Einstein's field equations. Without loss of generality we assume that the black holes M_1, M_2 are at the points $z = +1, z = -1$ respectively. These points represent the horizons of the two black holes.

In the case of photons we have $\delta_1 = 0$, and in the case of particles $\delta_1 = 1$. The potential U is the classical potential of two fixed centers (opposite to the potential energy), which is a well known case in Celestial Mechanics (Charlier 1902, Deprit 1960). The addition of 1 is irrelevant in classical mechanics, but it is necessary in relativity.

The classical problem is integrable. The equations of motion can be solved explicitly using elliptic integrals. Thus it came as a surprise that the relativistic problem is chaotic. More accurately the null geodesics are completely chaotic, while the time-like geodesics are partly chaotic.

We can write the Eulerian equations of motion, corresponding to the Lagrangian (1), as follows

$$p_t = \frac{\dot{t}}{U^2} = E \quad (5)$$

(energy),

$$p_\phi = U^2 \sinh^2 \psi \sin \theta \dot{\phi} = L_z \quad (6)$$

(angular momentum).

We write also the momenta

$$p_\psi = -\frac{\partial L}{\partial \dot{\psi}} = U^2 Q \dot{\psi} \quad (7)$$

$$p_\theta = -\frac{\partial L}{\partial \dot{\theta}} = U^2 Q \dot{\theta} \quad (8)$$

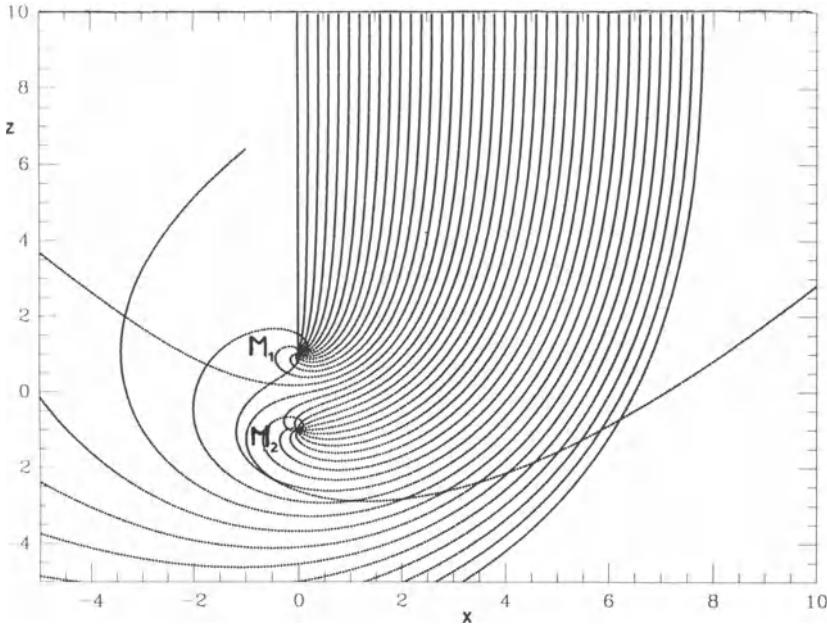


Figure 1. A set of initially parallel orbits coming from infinity (with $x_o > 0$) are separated into orbits falling into the black holes M_1 , or M_2 , and orbits escaping back to infinity.

where

$$Q = \cosh^2 \psi - \cos^2 \theta = \sinh^2 \psi + \sin^2 \theta > 0. \quad (9)$$

Then the equations of motion can be written in the form

$$\frac{d}{ds} \left(U^2 Q^2 \dot{\psi}^2 + \frac{L_z^2}{\sinh^2 \psi} \right) = E^2 \dot{\psi} \frac{\partial}{\partial \dot{\psi}} \left[U^2 Q \left(U^2 - \frac{\delta_1}{E^2} \right) \right] \quad (10)$$

$$\frac{d}{ds} \left(U^2 Q^2 \dot{\theta}^2 + \frac{L_z^2}{\sin^2 \theta} \right) = E^2 \dot{\theta} \frac{\partial}{\partial \dot{\theta}} \left[U^2 Q \left(U^2 - \frac{\delta_1}{E^2} \right) \right] \quad (11)$$

These equations are equivalent to those of Chandrasekhar (1989) but more concise. From now on we consider orbits on the meridian plane $\varphi = 0$, therefore

$$L_z = 0. \quad (12)$$

Then an integral of motion is

$$U^2 \left[E^2 - Q \left(\dot{\psi}^2 + \dot{\theta}^2 \right) \right] = \delta_1. \quad (13)$$

2. ORBITS OF PHOTONS (NULL GEODESICS)

If $\delta_1 = 0$ the orbits are the same for any value of E , therefore there is no loss of generality if we take $E = 1$.

If we consider a set of parallel orbits coming from infinity (Fig. 1) we see that they can be separated into three types:

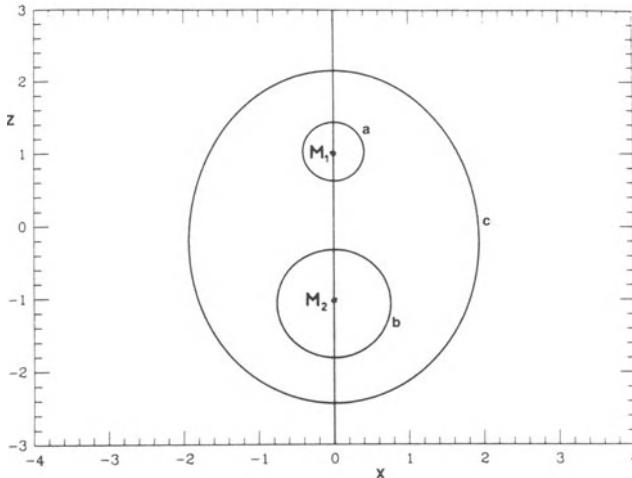


Figure 2. The main types of periodic orbits of photons around two black holes: (a) around M_1 , (b) around M_2 , and (c) around both M_1 and M_2 . ($M_1 = 0.5$, $M_2 = 1$).

- (I) Orbits falling into the black hole M_1 ,
- (II) Orbits falling into the black hole M_2 , or
- 'III) Orbits escaping the infinity.

We notice that the topology of the orbits is such, that

- (a) between two orbits of types I and II here is a set of orbits of type III.
- (b) between two orbits of types I and III there is a set of orbits of type II, and
- (c) between two orbits of types II and III there is a set of orbits of type I.

The only orbits that do not belong to the types (I), (II) and (III) are the periodic orbits, but all of them are unstable, and they do not trap any nonperiodic orbits around them.

We have found several families of periodic orbits and we checked that all of them are unstable. However we could not give an analytic proof of their instability. (On the contrary in the case of particle orbits we found both stable and unstable periodic orbits).

The periodic orbits play a crucial role in separating the orbits of type (I), (II) and (III). In fact the separating orbits are asymptotic to three basic unstable periodic orbits that we call (a), (b) and (c) (Fig. 2). The orbit (a) surrounds the black hole M_1 , the orbit (b) surrounds the black hole M_2 and the orbit (c) surrounds both black holes. The orbits (a) and (b) are called "satellite" orbits.

If an orbit crosses the orbits (a) or (b) inwards it falls into the black holes M_1 and M_2 respectively, and if it crosses the orbit (c) outwards it escapes to infinity. As

a consequence the set of orbits falling into a black hole, say M_2 , is limited by orbits approaching asymptotically the periodic orbit (b) clockwise or counterclockwise (Fig. 3). All orbits between these limiting orbits fall into this black hole. Similar asymptotic curves exist for the orbits (a) and (c). But if we start orbits a little outside the interval between the two asymptotic curves of Fig. 3 we have infinite sets of orbits of types (I), (II) and (III). The limits of these sets are three Cantor sets, that are mixed in a most intricate way.

These properties define what we call “chaotic scattering”, or “chaotic diffusion” (Eckhardt and Jung 1986, Jung and Scholz 1987, 1988, Hénon 1988, 1989, Bleher et al. 1990). Similar phenomena have been observed in dissociations of molecules (Noid and Koszykowski 1980, Woll and Hase 1980, Agmon 1982, Hedges and Reinhart 1983, Noid

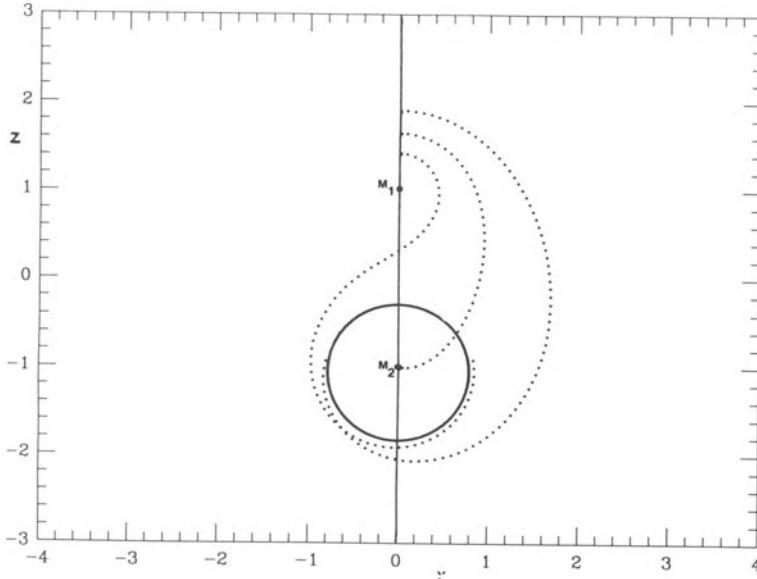


Figure 3. Asymptotic orbits to the periodic orbit (b) around M_2 . These orbits start perpendicularly to the z -axis above M_1 and approach the orbit (b) asymptotically clockwise, or counterclockwise. All orbits starting perpendicularly between these two orbits fall into the black hole M_2 . ($M_1 = 0.5$, $M_2 = 1$).

et al. 1986), in interactions of asteroids (Petit and Hénon 1986), or in stars escaping from galaxies (Contopoulos 1990b, Contopoulos and Kaufmann 1992, Contopoulos et al. 1993).

The initial conditions of the orbits of types (I), (II) and (III) are denumerable sets of intervals of positive measure, but their boundaries are accumulation points of boundaries of the same type. Some of these sets are relatively large, but other sets are extremely small. This is seen in Fig. 4, where a small beam of photons, coming from infinity, is split in a complicated way in sets of orbits of types (I), (II) and (III). In fact it is split into an infinity of sets of all three types, but orbits making several turns around M_1 , or M_2 , are not marked there.

All periodic orbits have in their neighborhood an infinity of sets of orbits of types

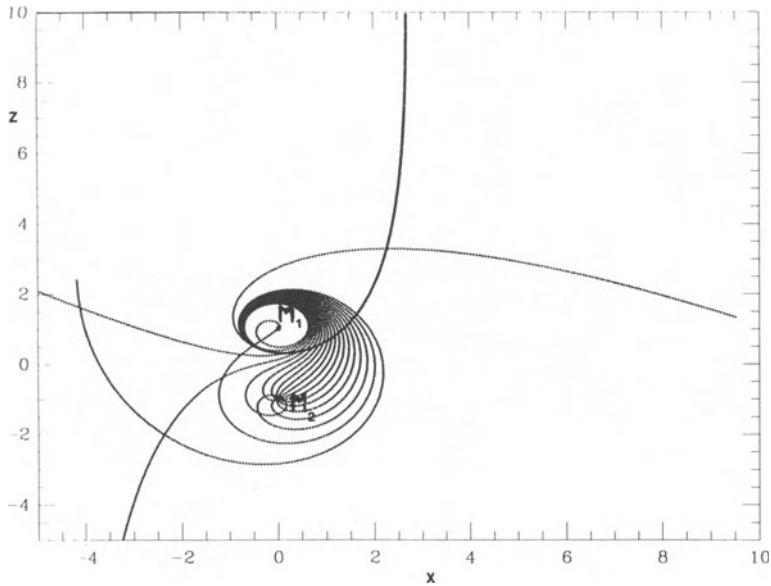


Figure 4. A thin beam of orbits of photons coming from infinity which is split into orbits of type (I) (falling into M_1), (II) (falling into M_2) and (III) (escaping to infinity).

(I), (II) and (III). These infinite sets exist on both sides of the periodic orbits, except for the periodic orbits (a), (b), (c), which have only on one side orbits of type (I), (II) and (III) respectively.

3. ORBITS OF PARTICLES (TIME-LIKE GEODESICS)

In the case of particles ($\delta_1 = 1$) the escape energy is $E = 1$. If $E > 1$ there are orbits going to infinity and the situation is similar to that of photons. In fact when $E \rightarrow \infty$ we tend to the case $\delta_1 = 0$ [Eqs. (10) and (11)].

But if $E < 1$ there is a curve of zero velocity (CZV) defined by Eq. (13) if $\dot{\psi} = \dot{\theta} = 0$, which limits the extent of the orbits. The equation of the CZV is

$$U = E^{-1} \quad (14)$$

For relatively large E the CZV is like an ellipse around both M_1 and M_2 . It is symmetric with respect to the z -axis, but asymmetric with respect to the x -axis, unless $M_1 = M_2$. For smaller E the CZV develops a “throat” between M_1 and M_2 close to the z -axis (Fig. 5), and for still smaller E it splits into two curves surrounding M_1 and M_2 separately. The transition to a double curve occurs when

$$0 < \varepsilon \equiv \frac{1}{E} - 1 = \varepsilon_1 = \frac{1}{2} \left(\sqrt{M_1} + \sqrt{M_2} \right)^2 \quad (15)$$

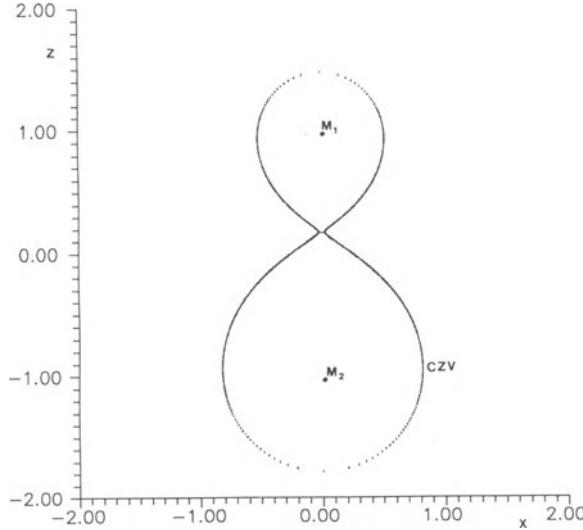


Figure 5. A curve of zero velocity (CZV) for $E = \sqrt{0.1657}$. This value is slightly larger than $E = E_1$ (when the CZV splits into two curves). The short curve at the throat of the CZV is a hyperbolic periodic orbit.

(Contopoulos and Papadaki 1993). Therefore if

$$1 > E > E_1 = \frac{2}{2 + (\sqrt{M_1 + M_2})^2}, \quad (16)$$

the CZVs surround both M_1 and M_2 , while if

$$E < E_1 \quad (17)$$

the CZVs are split into two curves surrounding M_1 , and M_2 separately. For $E = E_1$ the CZV has a figure eight form. Finally when

$$E = E_o = 0 \quad (18)$$

the CZV are reduced to the black holes M_1 and M_2 .

In the case of particle orbits with $E < 1$ we do not have orbits of type (III) (escaping to infinity). But we may have two new types of orbits:

(IV) Quasi periodic orbits close to a stable periodic orbit.

(V) Chaotic orbits that fill some region inside the CZV, but do not fall into M_1 , or M_2 .

Chaotic (or stochastic) orbits of type (V) are generated by some unstable periodic orbits, as we will see below.

The periodic orbits (a), (b) and (c) exist in the case of particles only for certain intervals of values of the energy and of the masses of the black holes. If we keep $M_2 = 1$

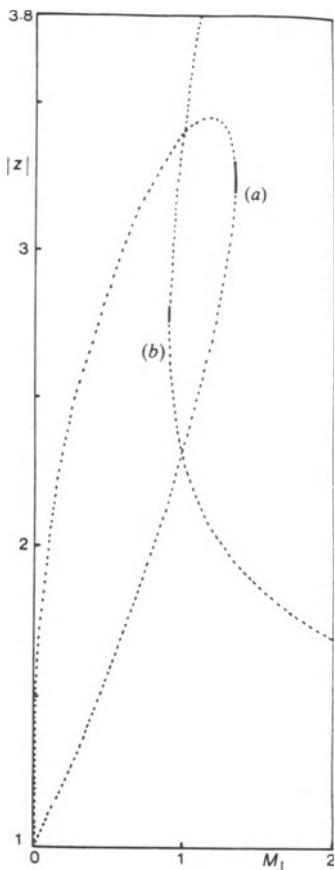


Figure 6. The characteristics of the families (a) and (b) for $E = \sqrt{0.5}$. The solid curves indicate stable orbits, and the dotted curves unstable orbits.

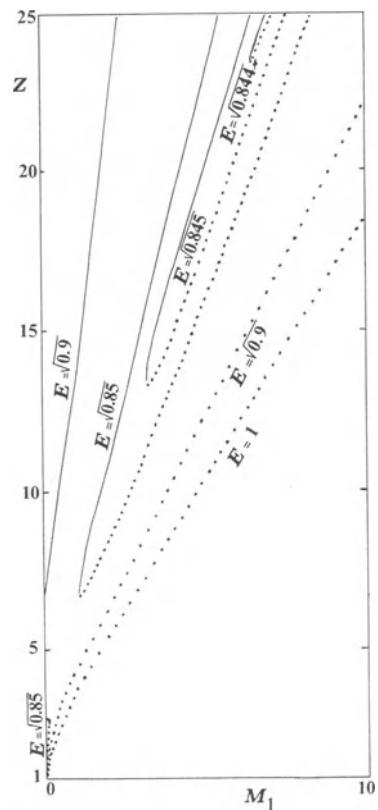


Figure 7. The characteristics of the family (c) for various values of E . Symbols as in Fig. 6.

we construct the characteristics of the families (a), (b) and (c) for various values of the energy E (Figs. 6 and 7). Namely we find the perpendicular intersections z of these orbits as functions of M_1 for fixed E . In the case of the families (a) and (c) we give the largest value of z ($z > 1$), while in the case of the family (b) we give the absolute value of the smallest value of z ($z < -1$).

In Figs. 6 and 7 we see the following:

- a) For every value of M_1 there are in general two or zero orbits of types (a), (b) and (c).
- b) The family (a) exists for small M_1 ($0 \leq M_1 \leq M_a = 1.32576$ for $E = \sqrt{0.5}$) but not for larger M_1 .
- c) The family (b) exists for large M_1 ($0.90614 = M_b \leq M_1$ for $E = \sqrt{0.5}$) but not for smaller M_1 .
- d) The family (c) does not exist at all for $E \lesssim \sqrt{0.8438}$. For somewhat larger E there are two intervals of M_1 for which this family exists. One is close to zero, up to a certain maximum $M_1 = M_{\max}$, and the other above a certain minimum $M_1 = M_{\min}$ (Fig. 7), extending probably to $M_1 \rightarrow \infty$. For $E \simeq \sqrt{0.855}$, M_{\max} and M_{\min} join and for larger E the orbits (c) exist for all the values of M_1 .

The upper characteristics of the family (c) are stable, while the lower characteristics are unstable (Fig. 7).

Orbits far outside the black holes M_1 and M_2 are trapped around the stable periodic orbit (c) surrounding the two black holes (orbits of type (IV); Fig. 8).

On the other hand the characteristics (a) and (b) (Fig. 6) have only small stable parts near their maximum respectively. In fact it is known (Poincaré 1892) that at the point where a characteristic reaches a maximum or a minimum we have the joining of a stable and an unstable family (this is called sometimes a “tangent bifurcation”). The same happens in Fig. 7. But while the upper branch of the family (c) is always stable, the upper branch of the stable family (a) becomes unstable, as M_1 decreases, by period doubling, and for still smaller M_1 we have an infinity of period doubling bifurcations (Fig. 9), that lead to chaos according to the scenario of Benettin et al. (1980), which is similar to the scenario found by Feigenbaum (1978) and by Coullet and Tresser (1978) for dissipative systems. The intervals between successive period doubling bifurcations decrease by a factor

$$\delta = 8.72, \quad (19)$$

which is the universal bifurcation ratio for conservative systems (while $\delta = 4.67$ for dissipative systems).

Besides the period doubling bifurcations there are also higher order bifurcations (Fig. 9). In particular the family 3 has an unstable branch close to 1 and a stable branch further away. On a Poincaré surface of section (z, \dot{z}) (i.e. for $x = 0$; Fig. 10) we have 3 islands of stability near the stable orbit 3 (orbits of type (IV)) and some chaos near the unstable orbit 3 (orbits of type (V)). But this chaotic region is separated from the large outer region, where the chaotic scattering is effective, by the outermost invariant curve of Fig. 10. This is a KAM curve (after Kolmogorov, Arnold and Moser; see Arnold 1978), which is an intersection of a torus separating the neighborhood of the orbit (a) from the external region, and does not allow communication between them. Thus, although we have chaotic orbits near the unstable periodic orbit of multiplicity 3, these orbits never reach the black holes M_1 , or M_2 .

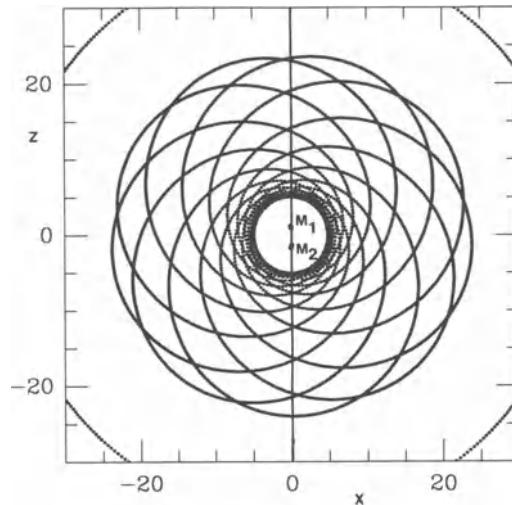


Figure 8. A quasiperiodic orbit far from the two black holes M_1, M_2 . (Case $M_1 = M_2 = 1, E = \sqrt{0.9}$ initial $z = 5, x = \dot{z} = 0$). The outer arcs belong to the CZV.

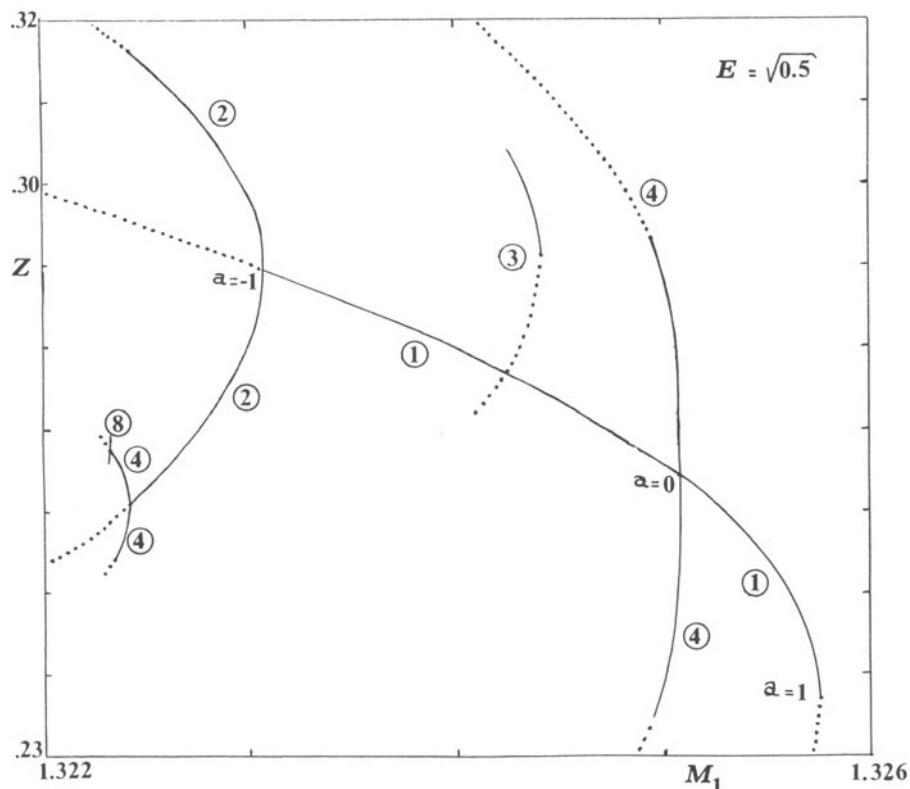


Figure 9. Characteristics of families of various multiplicities bifurcating from the stable family (a), marked by 1. Note the sequence of period doubling bifurcations $1 \rightarrow 2, 2 \rightarrow 4, 4 \rightarrow 8$. a is the Hénon (1965) stability parameter, (—) stable, (····) unstable orbits. $E = \sqrt{0.5}$.

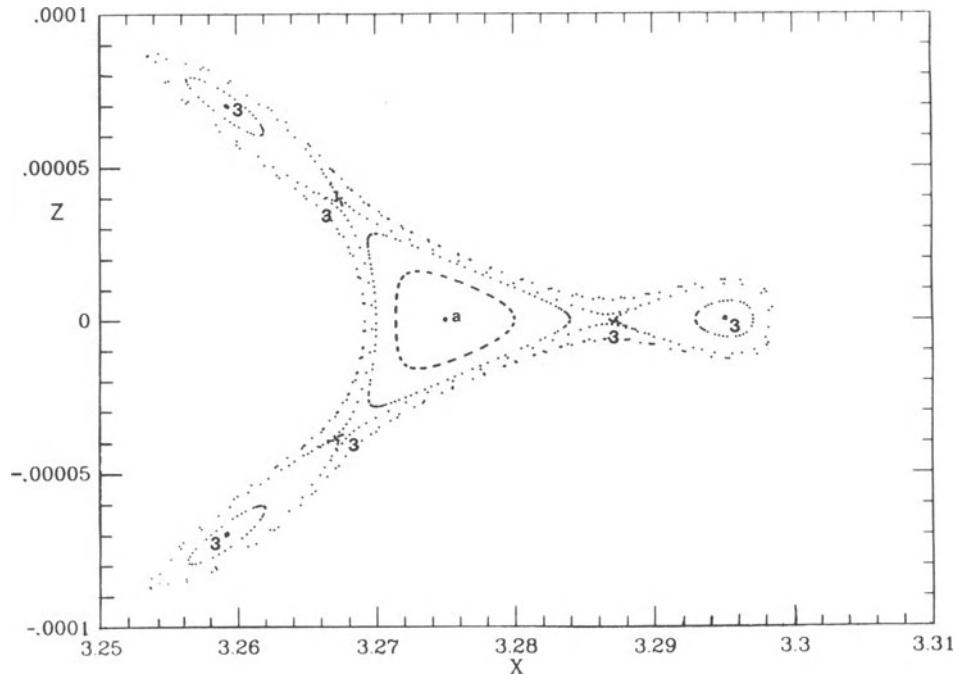


Figure 10. Invariant curves and chaos on a surface of section \$(z, \dot{z})\$ (\$x = 0\$). The stable periodic orbit (a) belongs to the upper branch of Fig. 6. \$M_1 = 1.3244\$, \$E = \sqrt{0.5}\$.

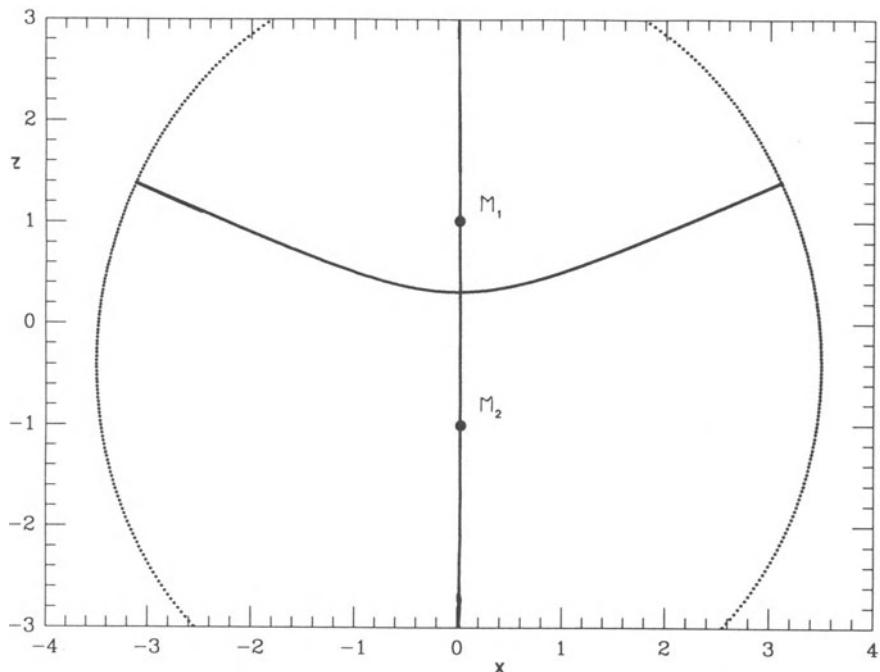


Figure 11. An unstable periodic orbit, like a hyperbola, reaching perpendicularly the CZV. \$E = \sqrt{0.5}\$, \$M_1 = 0.5\$, \$M_2 = 1\$, \$z = 0.31122\$.

On the other hand orbits starting outside the outermost invariant curve of Fig. 10 after some chaotic wanderings end up at one of the black holes M_1 , or M_2 .

We consider now the separation of the orbits reaching the black holes M_1 , or M_2 . If both orbits (a) and (b) exist then the inner orbits (a) and (b) (those closer to the center) play the same role as in the case of photons. Namely orbits crossing the inner orbit (a) fall into M_1 , and orbits crossing the inner orbit (b) fall into M_2 . If we consider a particular set of orbits (e.g. orbits crossing perpendicularly the z -axis M_1) the orbits falling into M_1 , or M_2 , are limited by asymptotic orbits, as in Fig. 3. The main difference is now that outside these two asymptotic curves there are only two Cantor sets limiting the sets of orbits falling into M_1 , or M_2 . The third set, of orbits escaping to infinity, does not exist any more. Furthermore the outer regions inside the curve of zero velocity may contain orbits trapped near stable periodic orbits, as in the case of Fig. 8 (one orbit in configuration space), or Fig. 10 (several orbits marked by their successive intersections (consequents) on a Poincaré surface of section). The boundary between the trapped orbits around the stable periodic orbits (of type (IV)) and the untrapped orbits of types (I), or (II) is not well defined. In such cases we know that the “last KAM curves” are fractals (see e.g. Greene 1969).

If one type of orbits (a), or (b), does not exist, then the separation of orbits (I) and (II) is realized in a different way (Contopoulos 1991). E.g. in the case $E = \sqrt{0.5}$, $M_1 = 0.5$, $M_2 = 1$, the orbit (b) does not exist. Then there is an unstable orbit like a hyperbola (Fig. 11), such that all orbits crossing it downwards fall into M_2 .

If we consider the set of orbits starting perpendicularly to the z -axis above M_1 , the limit of the orbits falling into M_2 is an orbit asymptotic to the “hyperbolic” orbit of Fig. 11.

The “hyperbolic” orbits exist when $E > E_1$. When E is slightly larger than E_1 the “hyperbolic” orbit bridges the throat of the CZV (Fig. 5), and tends to one point when $E \rightarrow E_1$. We have found that the family (a) does not exist for $E^2 < 0.375$. For $E^2 = 0.375$ the family (b) exists for $M_1 > 1.517$. Thus there is an interval of values of M_1 , in which there is neither an orbit (a) nor an orbit (b).

For values of $E^2 > 0.375$ the family (a) exists from $M_1 = 0$ to a maximum $M_1 = M_{\max}$ and the family (b) from a minimum $M_1 = M_{\min}$ to $M_1 \rightarrow \infty$.

For $0.375 < E^2 < 0.48$ there is an interval of values of M_1 , between M_{\max} and M_{\min} , in which neither (a) nor (b) exists. For $E^2 \geq 0.48$ we have $M_{\max} > M_{\min}$ therefore there is an interval of values of M_1 in which both (a) and (b) exist.

Finally for $E^2 \geq 0.8$ we have $M_{\min} = 0$, i.e. the family (b) exists for all $M_1 \geq 0$, while the family (a) exists from $M_1 = 0$, up to a large M_{\max} .

4. COMPARISON OF RELATIVISTIC AND NEWTONIAN ORBITS

The Newtonian analog of the problem of two black holes is found if we use the approximations

$$U^2 = 1 + \frac{2W}{Q}, \quad U^4 = 1 + \frac{4W}{Q}, \quad (20)$$

where

$$W = (M_1 + M_2) \cosh \psi + (M_1 - M_2) \cos \theta. \quad (21)$$

Then the equations of motion can be separated and give (for $\delta_1 = 1$) two integrals of motion:

$$\frac{1}{2}U^4Q^2\dot{\psi}^2 = H \cosh^2 \psi + m \cosh \psi + \alpha, \quad (22)$$

$$-\frac{1}{2}U^4Q^2\dot{\theta}^2 = H \cos^2 \theta - m' \cos \theta + \alpha, \quad (23)$$

where

$$H = \frac{1}{2}(E^2 - 1), \quad (24)$$

$$m = 2(M_1 + M_2) \left(E^2 - \frac{1}{2} \right), \quad m' = 2(M_1 - M_2) \left(E^2 - \frac{1}{2} \right), \quad (25)$$

and α is a constant of integration. If we set

$$\lambda = \cosh \psi, \quad \mu = -\cos \theta \quad (26)$$

we write the equations of motion in the classical form

$$\frac{d\lambda}{\sqrt{R(\lambda)}} = \frac{d\mu}{\sqrt{S(\mu)}} = dt, \quad (27)$$

where

$$R(\lambda) = 2(\lambda^2 - 1)(H\lambda^2 + m\lambda + \alpha), \quad (28)$$

$$S(\mu) = 2(\mu^2 - 1)(H\mu^2 + m'\mu + \alpha), \quad (29)$$

(Charlier 1902). These equations can be solved in terms of elliptic integrals.

The escape energy is $E = 1$, which corresponds to the classical energy $H = 0$. The CZVs can be defined for all $\sqrt{0.5} < E < 1$. The transition value at which the CZV is of figure eight form is (Contopoulos and Papadaki 1993)

$$E = E_1 = \left[\frac{2(M_1 - M_2)^2 + 3(M_1 + M_2) + 1 - 2\sqrt{M_1 M_2}}{4(M_1 - M_2)^2 + 4(M_1 + M_2) + 1} \right]^{\frac{1}{2}}. \quad (30)$$

For $E_1 < E < 1$, the CZV surrounds both points M_1 and M_2 .

The value of E_1 for $M_1 = 0.5$ and $M_2 = 1$ in the relativistic case is $E_1 = 0.40698$ while in the Newtonian case it is $E_1 = 0.7571151$.

We have calculated, in particular, several families of periodic orbits in the Newtonian case, and compared these orbits with the corresponding relativistic periodic orbits.

The most surprising result of our study was that the families of "satellite" orbits (a) and (b), that close after one revolution around M_1 and M_2 respectively, do not exist in the Newtonian case. This result, that escaped discovery, despite the detailed studies of the classical problem over several decades, was found first numerically, but it was also proved analytically (Appendix B of Contopoulos 1990). Only when $M_2 = 0$, or $M_1 = 0$ (exactly) there are Keplerian orbits around M_1 , or M_2 , closing after one revolution. On the other hand there are higher order orbits of various types around M_1 alone, or M_2 alone, or around both M_1 and M_2 . E.g. there are orbits like a figure eight, like the letter γ (gamma), or an inverse γ (Fig. 12), etc. There are also exactly elliptic orbits surrounding both black holes and exactly hyperbolic orbits reaching the CZV.

In the Newtonian case the orbits have more restrictions than in the relativistic case. The orbits are usually confined between an ellipse and a hyperbola. In the case of Fig. 12, where the CZV is separated into two closed curves there are two periodic orbits,

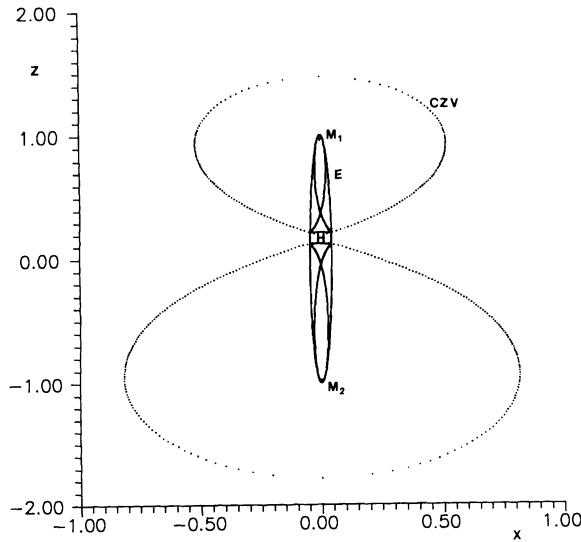


Figure 12. A gamma and an inverse gamma orbit in the Newtonian case, for $E^2 = 0.573223304$. The CZV is split into two curves. The orbits are confined inside a common ellipse E and two hyperbolas H .

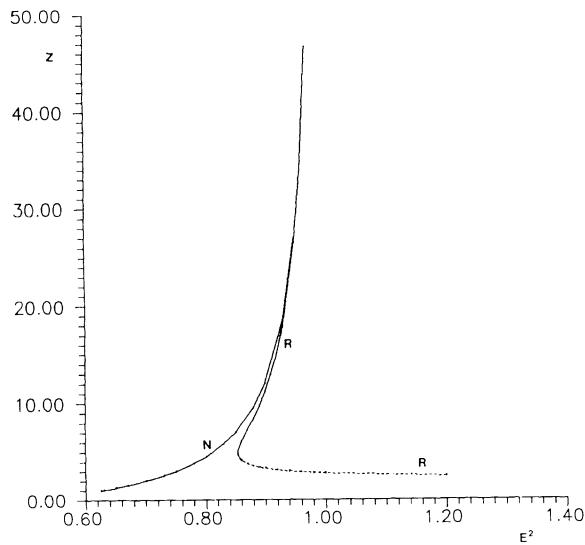


Figure 13. The Newtonian and the relativistic characteristics of the family (c) of elliptic-type orbits (N) and (R) around both black holes. The maximum z is given as a function for E^2 for $M_1 = M_2 = 1$. (—) stable, (····) unstable orbits.

one around M_1 (of inverse gamma type) and the other around M_2 (of type gamma). Each orbit is inside an ellipse E and a hyperbola H . The ellipse crosses the CZV, and surrounds both orbits while there are two limit hyperbolae reaching the CZV.

The number and size of the periodic orbits may be different in the Newtonian and the relativistic case. In Fig. 13 we give the characteristics of a Newtonian elliptical orbit and of the corresponding relativistic orbit. These orbits surround both M_1 and M_2 and intersect the z -axis twice perpendicularly.

We give the point of intersection of each orbit with the z -axis above M_1 as a function of E^2 . The Newtonian characteristic is a monotonic function. The relativistic characteristic practically coincides with the Newtonian one for large z (far from the black holes), but it is very different close to the black holes. It has a $E_{\min}^2 \simeq 0.625$, where it joins another branch, which is unstable. The new branch extends to values larger than 1. For $E > 1$ orbits that cross this orbit outwards escape to infinity. On the other hand the upper branch is stable and tends to $z \rightarrow \infty$ when $E \rightarrow 1$.

Thus the Newtonian and relativistic periodic orbits are similar far from the black holes, but they are very different close to the black holes. Therefore the relativistic case is not a small perturbation of the Newtonian case.

REFERENCES

- Agmon, N., 1982, *J. Chem. Phys.*, **76**, 1309.
 Arnold, V. J., 1978, *Mathematical Methods of Classical Mechanics*, Springer Verlag.
 Benettin, G., Cercignani, C., Galgani, L. and Giorgilli, A., 1980, *Lett. Nuovo Cim.*, **28**, 1.
 Bleher, S., Grebogi, C. and Ott, E., 1990, *Physica*, **D46**, 87.
 Chandrasekhar, S., 1989, *Proc. R. Soc. Lond.*, **A421**, 227.
 Charlier, C. L., 1902, *Die Mechanik des Himmels*, von Veit, Leipzig.
 Contopoulos, G., 1990a, *Proc. R. Soc. Lond.*, **A431**, 183.
 Contopoulos, G., 1990b, *Astron. Astrophys.*, **231**, 41.
 Contopoulos, G., 1991, *Proc. R. Soc. Lond.*, **A435**, 551.
 Contopoulos, G. and Kaufmann, D., 1992, *Astron. Astrophys.*, **253**, 379.
 Contopoulos, G., Kandrup, H.E. and Kaufmann, D., 1993, *Physica*, **D64**, 310.
 Contopoulos, G. and Papadaki, H., 1993, *Celest. Mech. Dyn. Astron.*, **55**, 47.
 Coullet, P. and Tresser, C., 1978, *J. Phys. Paris*, **C5**, 25.
 Deprit, A., 1960, *Mathematiques du XXe Siècle*, Univ. Louvain 1, 45.
 Eckhardt, B. and Yung, C., 1986, *J. Phys.*, **A19**, L829.
 Feigenbaum, M. J., 1978, *J. Stat. Phys.*, **19**, 25.
 Greene, J. M., 1969, *J. Math. Phys.*, **20**, 1183.
 Hedges, R. M. and Reinhart, W. P., 1983, *J. Chem. Phys.*, **78**, 3964.
 Hénon, M., 1965, *Ann. Astrophys.*, **28**, 992.
 Hénon, M., 1988, *Physica*, **D33**, 132.
 Hénon, M., 1989, *La Recherche*, **20**, 490.
 Majumdar, S. D., 1947, *Phys. Rev.*, **72**, 390.

- Noid, D. W. and Koszykowski, M. L., 1980, *Chem. Phys. Lett.*, **73**, 114.
- Noid, D. W., Gray, S.K. and Rice, S. A., 1986, *J. Chem. Phys.*, **84**, 2649.
- Papapetrou, A., 1947, *Proc. R. Irish Acad.*, **51**, 191.
- Petit, J. M. and Hénon, M., 1986, *Icarus*, **66**, 536.
- Poincaré, H., 1892, *Méthodes Nouvelles de la Mécanique Céleste*, Gauthier Villars, Paris.
- Woll, R. J. and Hase, W., 1980, *J. Chem. Phys.*, **73**, 3779.
- Yung, C. and Scholz, H.-J., 1987, *J. Phys.*, **A20**, 3607.
- Yung, C. and Scholz, H.-J., 1988, *J. Phys.*, **A22**, 2925.

PARTICLE MOTION AROUND PERTURBED BLACK HOLES: THE ONSET OF CHAOS

Luca Bombelli¹

RggR, Université Libre de Bruxelles,
Campus Plaine CP 231, 1050 Brussels, Belgium

Abstract. I describe an example of relativistic chaos that arises in the motion of a relativistic particle around a black hole. In the case of a static, Schwarzschild black hole, for which geodesic motion is integrable, there are, in addition to the stable circular orbits present in Newtonian gravity, unstable circular orbits near $r = 6M$. These give rise to other orbits, which approach the unstable ones for $t \rightarrow \pm\infty$ and wind outwards up to some maximum radius in between. The latter are examples of the so-called homoclinic orbits in the phase space of an integrable system, which are often seeds of chaos when a perturbation is added to the system. By using the Melnikov method for detecting homoclinic chaos in near integrable systems, we conclude that, as one adds a time-periodic perturbation to the static black hole, a region of chaotic motion replaces the homoclinic orbit.

1. INTRODUCTION

The highly non-linear nature of general relativity theory has, as has often been emphasized, a number of practical consequences. One of these is that relatively few exact solutions of the gravitational field equations are known, and they are special in the sense that they are always obtained by introducing at least some degree of symmetry in the equations; a second consequence is that the behavior of the solutions may be unstable against small perturbations. This gives rise to a situation in which, on the one hand, observational predictions of the theory are based on the few known solutions, and, on

¹e-mail: lbombell@ulb.ac.be

the other hand, there are good reasons to believe that those solutions may not be representative of all the qualitative features present in the more realistic situations they are supposed to approximate. Furthermore, the same can be said about the motion of small bodies, considered as test particles, in a gravitational field, since their behavior may be unstable under perturbations of the background metric. This situation is being gradually modified by numerical simulations carried out with increasingly sophisticated methods, but still few simulations tackle problems that are fully 3-dimensional, i.e., without symmetries, and the use of numerical methods in fact raises new concerns precisely with respect to one of the features of the theory one would like to probe: the presence of chaos.

As testified by the variety of papers in this volume, chaotic behavior has been studied in several general relativistic systems, both in the evolution of the metric itself, in cosmological settings, and in test particle motion in a gravitational field. As a result, despite a controversy over whether certain cosmological models are chaotic or not, which has in part evolved into a discussion on the relationship between mathematical definitions of chaos and a physical notion of dependence on initial conditions and unpredictability, the existence of chaos in these systems is now widely accepted. The instability inherent in chaos however causes numerical work in this area to be sometimes viewed with suspicion. The finite precision of computers can bring in errors that act as noise affecting the dynamics, and possibly changing the character of the evolution. While in certain cases one might actually want to introduce some noise or perturbation, precisely to test the stability of a system, one would then like to have some control over the type of perturbation, and a theoretical model to describe its effect.

At any rate, with relativistic chaos relatively well established, one can start to go beyond just finding chaos in a system, and deal with the issue of its consequences for observational predictions. It is with this idea in mind that I started to collaborate with E. Calzetta, with the goal of looking for examples of relativistic systems with chaotic features that could have an effect on observation, possibly by being related to current or planned test of general relativity. In addition, we were interested in examples which can be treated by analytical techniques, so that one need not trust a result obtained just using numerical simulations.

The example we chose to start with is that of a particle moving around a black hole. Test particle, i.e. geodesic, motion around a Schwarzschild or Kerr black hole is integrable, as is well known. However, if we want to model a real black hole with a smaller body moving around it, a useful system to test the predictions of general relativity because it involves a strong gravitational field, a couple of facts must be taken into account: no black hole, even classically, is exactly stationary, and no object around it will move exactly along a geodesic. In the work described here, our aim was to analyze whether there are qualitative differences between motion around a stationary and a perturbed black hole, the question of stability versus chaos. We thus studied geodesic motion around a black hole subject to a time-dependent periodic perturbation in the metric, and we showed that at least a region of phase space arises in which motion is chaotic (Bombelli and Calzetta 1992); a similar system was studied by Moeckel (1992). We neglected instead the emission of gravitational radiation by the “particle,” which has important consequences both for the qualitative features of the motion, and for the possibility of observing the effect, so the present result should not be considered yet as observationally relevant; work is continuing to overcome this limitation. However, the work described here represents a first step in that direction, and it illustrates the use of

the Melnikov method for the detection of homoclinic chaos in near integrable systems, which is a simple, analytic tool that can be applied also to other problems in relativistic dynamics (Calzetta and El Hasi 1993, and Calzetta, these Proceedings).

The Melnikov method is a technique for detecting chaos arising when a time-periodic perturbation is added to an integrable system which possesses a homoclinic orbit in phase space (Guckenheimer and Holmes 1983; Wiggins 1988; Calzetta, these Proceedings). More specifically, consider a one-dimensional Hamiltonian system (the method can be applied in higher dimensions as well) with an unstable equilibrium point $x = x_{\text{un}}$ in the potential $U(x)$. There are then trajectories which approach x_{un} as $t \rightarrow \pm\infty$, and, if there are other points at which the potential has a higher value than $U(x_{\text{un}})$, there is also at least one trajectory ℓ asymptotically approaching x_{un} for both $t \rightarrow -\infty$ and $t \rightarrow +\infty$; this is called a *homoclinic orbit*, and it is a separatrix or boundary between those orbits which are confined to a compact region of phase space on one side of the hyperbolic point and those which aren't. Homoclinic orbits are likely seeds of chaos when a time-periodic perturbation is added to the Hamiltonian. If

$$\begin{aligned} H(p, q, t) &= H_0(p, q) + \epsilon G(p, q, t - t_0) \\ G(p, q, t - t_0) &= \text{Re } G(p, q) e^{i\omega(t-t_0)}, \end{aligned} \quad (1)$$

where t_0 is the initial phase of the perturbation, the criterion for establishing whether a stochastic region will form near ℓ is to evaluate the Melnikov integral

$$I(t_0) = \int_{-\infty}^{\infty} dt \{H_0(p(t), q(t)), G(p(t), q(t), t - t_0)\}, \quad (2)$$

taken along the unperturbed homoclinic loop ℓ , where $\{\cdot, \cdot\}$ stands for the Poisson bracket; if $I(t_0)$ has transverse zeroes at discrete values of t_0 , then chaos does develop.

The result we obtained in our application may be an indication that test particle motion in a gravitational field is generically chaotic, and in fact several other studies have already found geodesic chaos in curved spacetime. This should not surprise us, if we think of how easy it is to produce chaotic motion in somewhat different contexts; for example, geodesic flux in a compact space of negative curvature or in a billiard with any convex section in its boundary is known to be chaotic. But even in relativity, more or less physically motivated examples are known, like motion around two fixed black holes (Contopoulos 1990 a, b), and in some cosmological spacetimes (Lockhardt *et al* 1982, Tomaschitz 1991) as well as in the Ernst spacetime (Karas and Vokrouhlický 1992); similarly, a study of chaos in the string capture by a black hole has been carried out (Larsen 1993). Studies of this type, applied to null geodesics in cosmological spacetimes, may affect our models for extracting information on the structure of the universe from the radiation we observe. Similarly, if after gravitational radiation is taken into account in the motion around the black hole, some form of chaotic behavior survives the dissipative effects of radiation, we may have to take this fact into account when using the radiation emitted by a binary system to infer the values of some parameters of the system. Furthermore, in view of the difficulties encountered even for minisuperspace models in defining covariantly chaos in the evolution of the metric, the existence of geodesic chaos for some congruence of geodesics has been proposed as a covariant definition of chaos in the underlying metric (Núñez-Yépez *et al* 1993); the remarks above would translate then into an indication of the pervasiveness of chaos in the gravitational field itself.

Section 2 of this paper discusses some aspects of geodesic motion around a Schwarzschild black hole that we will need for the analysis of the perturbed case; this analysis is performed in section 3, where we show using Melnikov's method that homoclinic chaos arises under the perturbations we consider. Additional remarks, in particular on possible developments, are contained in the concluding section 4.

2. MOTION AROUND THE UNPERTURBED BLACK HOLE

In this section, we begin by studying geodesic motion around the unperturbed Schwarzschild black hole, described by the metric

$$g_{ab} dx^a dx^b = -f dt^2 + f^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\phi^2), \quad (3)$$

where $f(r) := 1 - 2M/r$. (This part of the problem is of course a standard one—see, e.g., Wald (1984)—but we will need to study one particular set of orbits in more detail than is usually done.)

For a particle of mass μ on a curved background, a possible form for the action is

$$S = \int \mathcal{L} ds = \frac{\mu}{2} \int g_{ab} \dot{x}^a \dot{x}^b ds, \quad (4)$$

where s is the proper time along the world-line $x^\mu(s)$, and an overdot denotes differentiation with respect to s , with the constraint

$$g^{ab} p_a p_b = -\mu^2, \quad (5)$$

where as usual we have defined $p_a := \partial \mathcal{L} / \partial \dot{x}^a = \mu g_{ab} \dot{x}^b$. In the case of the Schwarzschild metric, the momenta $p_t = -H_0$ and p_ϕ are constants of the motion, and we set them to the fixed values $-E$ and L , respectively. These two first integrals of the motion, together with the Hamiltonian corresponding to the action (4), which generates s -evolution, or equivalently the constraint (5), take the form

$$\begin{aligned} \mu f \frac{dt}{ds} &= E \\ \mu r^2 \sin^2 \theta \frac{d\phi}{ds} &= L \\ \mu^2 \left(\frac{dr}{ds} \right)^2 + f(r) \left[\mu^2 + \mu^2 r^2 \left(\frac{d\theta}{ds} \right)^2 + \frac{L^2}{r^2 \sin^2 \theta} \right] &= E^2. \end{aligned} \quad (6)$$

Restricting ourselves as usual, without loss of generality, to orbits on the equatorial plane $\theta = \pi/2$, we obtain for the radial motion the one-dimensional problem of a particle moving in the potential

$$U(r) = -\frac{2M\mu^2}{r} + \frac{L^2}{r^2} - \frac{2ML^2}{r^3}. \quad (7)$$

In the present case, in order for the homoclinic orbit required by the Melnikov method to exist, $U(r)$ must have an unstable equilibrium point r_{un} at which its value is less than zero, the asymptotic value at $r = \infty$; there will then exist trajectories which

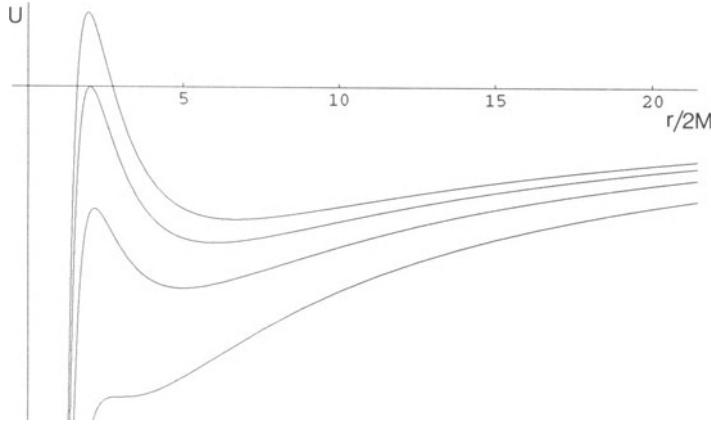


Figure 1. The potential U plotted as a function of $r/2M$ for $\beta = 0.55, 0.5, 0.4$, and 0.0 , from top to bottom.

approach r_{un} asymptotically as $t \rightarrow \pm\infty$, and bounce off the large r tail of the potential. The existence of such points is better discussed in terms of the parameter

$$\beta := \sqrt{1 - 12M^2\mu^2/L^2}. \quad (8)$$

The potential is plotted in figure 1 for different parameter values; it has a local maximum (and a minimum) for β real and positive, and its value there is negative if $\beta < \frac{1}{2}$; in other words, we must have

$$2\sqrt{3}M < \frac{L}{\mu} < 4M. \quad (9)$$

Notice that, had the general relativistic, r^{-3} term been absent in (7), there would always have been stable circular orbits for $L \neq 0$, as in the Newtonian theory, but no unstable ones, and thus no homoclinic orbit.

The explicit form for the homoclinic orbit can be found by solving the equations (6), in which this particular orbit is picked out by setting its energy to be $U(r_{\text{un}})$, with $dr/ds = 0$ at the unstable point. To solve the radial equation, it is convenient to use an inverse radial coordinate $x := 2M/r$; we shall reexpress the solution in terms of r , but parametrized by

$$z := \exp(\sqrt{\beta}\phi), \quad (10)$$

rather than s or t , which gives

$$r(z) = \frac{6M}{(1 - 2\beta + 3\beta[(z - 1)/(z + 1)])^2}. \quad (11)$$

A homoclinic orbit in physical (configuration) space, corresponding to $\beta = 0.4$, is shown in figure 2. To find its shape in phase space, necessary for later calculations, we also need to calculate $p_r = \mu f^{-1}dr/ds$, which gives

$$p_r(z) = \pm \frac{1}{f(r(z))} \left[E^2 - f(r(z)) \left(\mu^2 + \frac{L^2}{r^2(z)} \right) \right]^{1/2}, \quad (12)$$

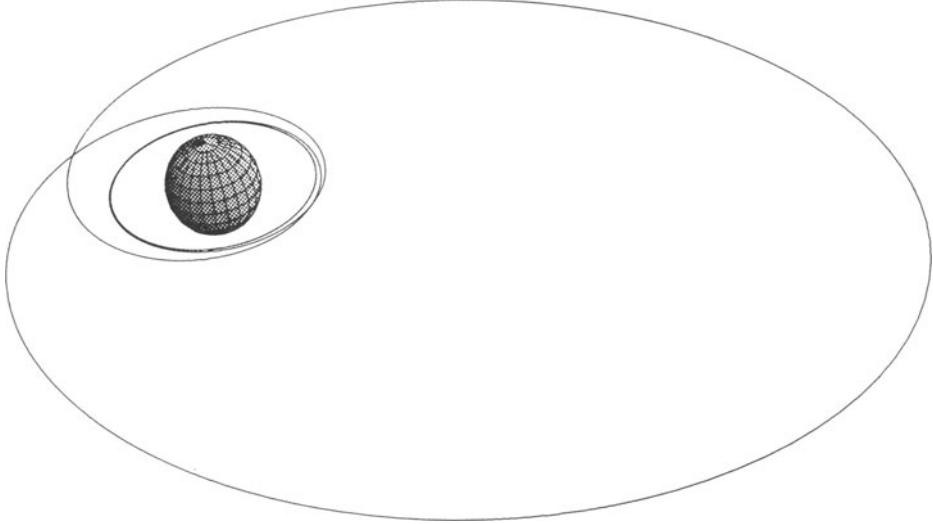


Figure 2. The homoclinic orbit around the unperturbed black hole for $\beta = 0.4$; the radial coordinate is the Schwarzschild coordinate r .

where the two signs correspond to the outgoing and ingoing parts of the orbit, respectively. Finally, to relate $r(z)$ and $p_r(z)$ to the time dependence of these quantities, we can combine and integrate the first two equations in (6), to get

$$t(z) = \frac{6\sqrt{6}(2-\beta)M}{\sqrt{\beta}(1+\beta)^{3/2}} \left[\frac{\ln z}{2-\beta} - \frac{2\beta}{1-2\beta} \frac{(1-5\beta)z + (1+\beta)}{(1+\beta)z^2 + 2(1-5\beta)z + (1+\beta)} + \right. \\ - \frac{\sqrt{\beta}(1+\beta)^{3/2}}{3\sqrt{6}(2-\beta)} \ln \frac{(2-\beta)z + (2+5\beta) - 2\sqrt{6\beta(1+\beta)}}{(2-\beta)z + (2+5\beta) + 2\sqrt{6\beta(1+\beta)}} + \\ \left. + \frac{\sqrt{\beta}(11-11\beta-4\beta^2)}{3\sqrt{3}(1-2\beta)^{3/2}} \arctan \frac{(1+\beta)z + (1-5\beta)}{2\sqrt{3\beta(1-2\beta)}} \right], \quad (13)$$

up to an additive constant, which is a real number and therefore it will not be important for us, as we shall see below.

3. MOTION AROUND THE PERTURBED BLACK HOLE

We now want to add a time-periodic perturbation to the Schwarzschild metric. For convenience, we will use only perturbations which are ϕ -independent and even under reflection with respect to the $\theta = \pi/2$ plane. This implies that L will still be conserved, the perturbed trajectory will stay in the equatorial plane, and we can still analyse the motion using the (r, p_r) phase space; the analysis could be extended however to arbitrary angular dependences. Components of the metric perturbation can be decomposed into tensor spherical harmonics labelled by (l, m) on the $r = \text{constant}$ spheres (Regge and Wheeler 1957). Then the requirements above translate into the condition $m = 0$, and restrictions on the values of l . Specifically, if we call even (odd) a tensor spherical harmonic whose sign changes by $(-1)^l$ ($(-1)^{l+1}$) under spatial inversion, the most

general such perturbation in the Regge-Wheeler gauge (Regge and Wheeler 1957) can be written as (the real part of)

$$h_{ab}(t, r, \theta) = \begin{bmatrix} 0 & 0 & 0 & h_0(r) \\ 0 & 0 & 0 & h_1(r) \\ 0 & 0 & 0 & 0 \\ h_0(r) & h_1(r) & 0 & 0 \end{bmatrix} e^{i\omega(t-t_0)} \sin \theta \frac{d}{d\theta} P_l(\cos \theta), \quad (14)$$

where P_l is the Legendre polynomial of order l , and we are interested in odd l (and real ω), while the most general even perturbation can be written in the form

$$h_{ab}(t, r, \theta) = \begin{bmatrix} fk_0(r) & k_1(r) & 0 & 0 \\ k_1(r) & f^{-1}k_0(r) & 0 & 0 \\ 0 & 0 & r^2 k_2(r) & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \theta k_2(r) \end{bmatrix} e^{i\omega(t-t_0)} P_l(\cos \theta), \quad (15)$$

where we are interested in even l . The dynamical content of the perturbation is thus reduced to the functions h_0 and h_1 , or k_0 , k_1 and k_2 , to be determined from the field equations.

The most reasonable and simplest perturbation of the form above we can consider is a solution of the linearized vacuum Einstein equation. Solving this equation, even for a simple background such as Schwarzschild, is a difficult task to which a considerable amount of literature has been devoted, but fortunately we only need to know the behavior of the solutions for large frequency ω . Then we obtain

$$\begin{aligned} h_0 &= A \left(\pm r + \frac{f}{i\omega} + \mathcal{O}(\omega^{-2}) \right) e^{\pm i\omega r^*} \\ h_1 &= A \left(\frac{r}{f} + \mathcal{O}(\omega^{-2}) \right) e^{\pm i\omega r^*}, \end{aligned} \quad (16)$$

where A is some possibly ω -dependent amplitude, for the functions appearing in the odd perturbation, and

$$\begin{aligned} k_0 &= B (1 + \mathcal{O}(\omega^{-1})) e^{\pm i\omega r^*} \\ k_1 &= \mp iB (1 + \mathcal{O}(\omega^{-1})) e^{\pm i\omega r^*} \\ k_2 &= \pm iB \left(\frac{f}{\omega r} + \mathcal{O}(\omega^{-2}) \right) e^{\pm i\omega r^*}, \end{aligned} \quad (17)$$

where B is some possibly ω -dependent amplitude, for the even perturbations. Here we have set as usual

$$r^* := r + 2M \ln \left(\frac{r}{2M} - 1 \right), \quad (18)$$

for the modified radial coordinate going to $-\infty$ at the horizon.

To calculate the Melnikov function (2), we need to know the Hamiltonian perturbation, $G(p, q, t - t_0)$. This is obtained by solving the mass shell constraint (5) with the perturbed metric,

$$f^{-1}(1 + \epsilon f h^{tt}) p_t^2 + 2\epsilon h^{tt} p_t p_t - (g^{ij} - \epsilon h^{ij}) p_i p_j = \mu^2, \quad (19)$$

for $-p_t = H_0 + \epsilon G$, having set $p_\theta = 0$, $p_\phi = L$. We obtain

$$H_0 = \sqrt{f(g^{ij} p_i p_j + \mu^2)} = \sqrt{f(fp_r^2 + \mu^2 + L^2/r^2)}$$

$$\begin{aligned}
G &= f h^{ti} p_i - \frac{f}{2 H_0} (h^{tt} H_0^2 + h^{ij} p_i p_j) \\
&= -f p_r h_{tr} - L r^{-2} h_{t\phi} - \frac{1}{2} H_0 f^{-1} h_{tt} + \\
&\quad - \frac{f}{2 H_0} (f^2 p_r^2 h_{rr} + 2 L f p_r r^{-2} h_{r\phi} + L^2 r^{-4} h_{\phi\phi}) . \tag{20}
\end{aligned}$$

The calculation of the Poisson brackets then gives, changing the variable of integration,

$$\begin{aligned}
I(t_0) &= \int_0^\infty \left\{ \left[\frac{2 f p_r}{r} h_{t\phi} - f p_r h_{t\phi,r} + \frac{1}{E r^2} \left(M E^2 - \frac{L^2 f^2}{r} - 7 M f^2 p_r^2 + 2 f^2 p_r^2 r \right) h_{r\phi} \right] + \right. \\
&\quad + \left[\frac{E M p_r}{L f} h_{tt} - \frac{p_r r^2}{2 L} h_{tt,r} + \frac{1}{L f} \left(M E^2 - \frac{L^2 f^2}{r} - M f^2 p_r^2 \right) h_{tr} + \right. \\
&\quad \left. - \frac{f^2 p_r^2 r^2}{L} h_{tr,r} + \frac{f p_r}{E L} \left(M E^2 - \frac{L^2 f^2}{r} - 2 M f^2 p_r^2 \right) h_{rr} + \right. \\
&\quad \left. - \frac{f^4 p_r^3 r^2}{2 E L} h_{rr,r} + \frac{L f p_r}{E r^3} \left(2 f - \frac{M}{r} \right) h_{\phi\phi} - \frac{L f^2 p_r}{2 E r^2} h_{\phi\phi,r} \right] \} \frac{dz}{\sqrt{\beta z}} , \tag{21}
\end{aligned}$$

where the orbit equations $r(z)$ and $p_r(z)$ are given by (11) and (12) respectively, and the perturbation h_{ab} by (14)–(17). This integral most likely couldn't be evaluated exactly even if we knew the exact form of h_{ab} . However, in the high frequency approximation, in which we do know h_{ab} , we can calculate $I(t_0)$ by the saddle point method.

All h_{ab} 's and their derivatives contain a factor $\exp\{i\omega(t \pm r^*)\}$. Let us restrict ourselves to the lower sign, and define $u := t - r^*$; the results for the upper sign would be analogous. Then our integral is of the form

$$I(t_0) \sim \operatorname{Re} e^{-i\omega t_0} \int_0^\infty dz e^{i\omega u(z)} F(r(z), p_r(z), \omega) , \tag{22}$$

where, for high ω , F is a slowly varying function of z with respect to the exponential, except at the singularities of F . Thus, if we deform the path of integration into the complex z -plane, making it pass through a critical point of u , the leading order contribution to the integral will come from near this critical point, provided that nowhere along the path is $\operatorname{Im} u(z) < 0$, so that the integrand stays finite as $\omega \rightarrow \infty$.

In addition to values of z corresponding to $r = 0$ and $r = 2M$, which are singularities of F , u has critical points at

$$z_\pm(\beta) = \frac{-2(1-2\beta) \pm 3i\sqrt{2\beta(1-\beta)}}{2(1+\beta) + \sqrt{6\beta(1+\beta)}} , \tag{23}$$

and it only has poles at $r = 0$ and $r = 2M$, corresponding to negative real values of z . In addition, we find that $\operatorname{Im} u(z)$ is positive at $z_+(\beta)$ and negative at $z_-(\beta)$, for all β . In fact, from plots of $\operatorname{Im} u(z)$ in the complex z -plane (an example of which, for $\beta = 0.4$, is shown in figure 3) we see that for any β the positive real axis can be deformed into a path through $z_+(\beta)$, which is always in the half plane $\operatorname{Im} z > 0$, without crossing a pole of F , and such that $\operatorname{Im} u(z) > 0$ all along the path.

The above shows that the saddle point method can be used to calculate the leading order contribution to $I(t_0)$, using $z = z_+(\beta)$ as the saddle point. Since, in addition, the function F (and its second derivative) can be evaluated at $z_+(\beta)$ and shown not to vanish there, the integral in (22) is not zero, and $I(t_0)$ vanishes only at the isolated,

transverse zeroes of $\exp(-i\omega t_0)$, at least for high enough ω . It seems reasonable to conclude then that $I(t_0)$ in fact vanishes identically at most for a discrete set of values of ω , if we assume it is an analytic function of ω . This is hardly surprising. Given the form of the perturbation, the only alternative would have been for $I(t_0)$ to be always identically zero; but this would have meant that all terms in the expansion of the integral in inverse powers of ω had to vanish, which they are unlikely to do unless there was some simple reason for it, e.g., as a consequence of the linearized Einstein equation. We conclude that for a generic vacuum perturbation h_{ab} of the type considered here, a region of phase space near the homoclinic orbit becomes chaotic.

4. CONCLUDING REMARKS

What the calculation in the previous two sections shows is that, when a black hole

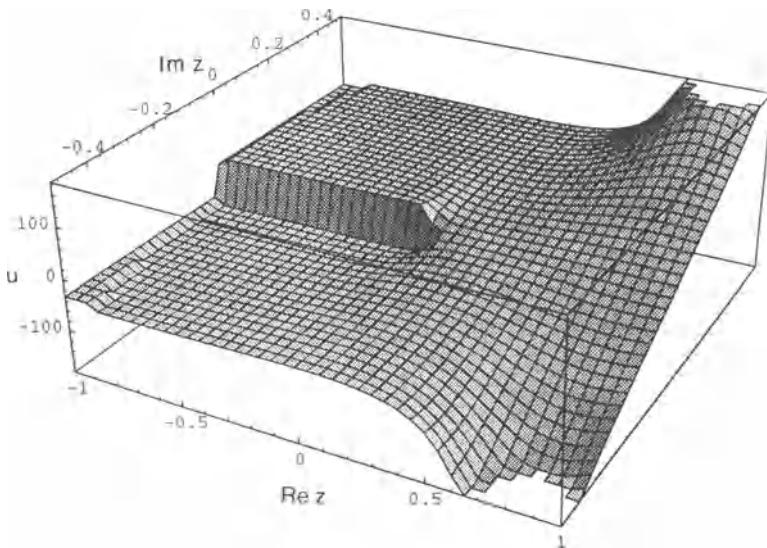


Figure 3. The imaginary part of $u(z)$ in the region $-1 < \text{Re } z < 1$, $-0.5 < \text{Im } z_0 < 0.5$, for $\beta = 0.4$. A branch cut from $z = 0$ in the negative real direction is visible.

metric is perturbed slightly away from a Schwarzschild solution by a time-periodic vacuum perturbation, the homoclinic orbits of the unperturbed, integrable system are replaced by stochastic layers where motion is chaotic. Natural questions that arise then concern the size of this region, and what happens to a test particle in one of the chaotic orbits; e.g., Does it fall into the hole? What happens when the perturbation does take the orbit out of the equatorial plane? While there exist techniques for estimating widths of stochastic regions and a version of the Melnikov method applicable when there are more than one degrees of freedom, it would be worthwhile to follow numerically various chaotic trajectories to get a clearer idea of what goes on.

There are various ways to make our model more realistic. One possibility would be to add a fluid to the exterior of the black hole, and consider the test particle in

the example discussed here as a model for a particle in an accretion disk. But cleaner, more precise predictions are obtained for a binary system of a black hole and a lighter companion; the main difference with respect to the work described here is that the system will lose energy by emitting gravitational radiation, and the lighter object will spiral towards the black hole. One consequence is that the lifetime of the system becomes finite, so that technically no chaos can be present, and some instabilities may be dissipated away. But the emitted radiation is in principle observable, and one can study the effect of any remaining instability of such a system under a perturbation in the metric on the observed radiation; this is the line along which work is continuing.

ACKNOWLEDGEMENTS

I would like to thank Esteban Calzetta and Mary Sakellariadou for discussions. This work was supported by the Directorate-General for Science, Research and Development of the Commission of the European Communities under contract no. CI1*-0540-M(TT).

REFERENCES

- [1] Bombelli, L. and Calzetta, E., 1992, "Chaos around a black hole", *Class. Quantum Grav.* **9**, 2573-99.
- [2] Calzetta, C., "Homoclinic chaos in relativistic cosmology", these Proceedings.
- [3] Calzetta, C. and El Hasi, C., 1993, "Chaotic Friedmann-Robertson-Walker cosmology", *Class. Quantum Grav.* **10**, 1825-41.
- [4] Contopoulos, G., 1990a, "Periodic orbits and chaos around two fixed black holes. I", *Proc. R. Soc. A* **431**, 183;
- [5] Contopoulos, G., 1990b, "Periodic orbits and chaos around two fixed black holes. II", *Proc. R. Soc. A* **435**, 551-62.
- [6] Guckenheimer, J. and Holmes, P., 1983, *Non-Linear Oscillations, Dynamical systems, and Bifurcations of Vector Fields*, (Berlin: Springer-Verlag).
- [7] Karas, V. and Vokrouhlický, D., 1992, "Chaotic motion in the Ernst space-time", *Gen. Rel. Grav.* **24**, 729-43.
- [8] Larsen, A.L., 1993, "Chaotic string capture by black hole", preprint hep-th/9309086.
- [9] Lockhart, C.M., Misra , B. and Prigogine, I., 1982, "Geodesic instability and internal time in relativistic cosmology", *Phys. Rev. D* **25**, 921-9.
- [10] Moeckel, R., 1992, "A nonintegrable model in general relativity", *Commun. Math. Phys.* **150**, 415-30.
- [11] Núñez-Yépez, H.N., Salas-Brito, A.L. and Sussman, R.A., 1993, "Geodesically chaotic spacetimes", preprint.
- [12] Regge, T. and Wheeler, J.A., 1957, "Stability of a Schwarzschild singularity", *Phys. Rev.* **108**, 1063-9.
- [13] Tomaschitz, R., 1991, "Relativistic quantum chaos in Robertson-Walker cosmologies", *J. Math. Phys.* **32**, 2571-9
- [14] Wald, R.M., 1984, *General Relativity*, (Chicago: University of Chicago Press).
- [15] Wiggins, S., 1988, *Global Bifurcations and Chaos*, (Heidelberg: Springer-Verlag).

CRITICAL BEHAVIOUR IN SCALAR FIELD COLLAPSE

M. W. Choptuik

Center for Relativity, Dept. of Physics
The University of Texas at Austin, Austin, TX 78712-1081

Abstract. I present an account of the discovery of critical behaviour in spherically-symmetric general-relativistic collapse of a scalar field, ϕ . Using an adaptive mesh-refinement algorithm in conjunction with finite-difference techniques, I have studied the non-linear evolution of parameterized initial configurations $\phi(r, 0; p)$, where a critical parameter value, $p = p^*$, generically demarcates the transition from spacetimes which do not contain a final black hole to spacetimes which do. The near critical regime, $p \approx p^*$, is characterized by a variety of non-linear phenomena including exponential sensitivity to initial conditions, scale-periodicity, and universal power-law dependence of black hole mass on parameter-space displacement $|p - p^*|$.

1. INTRODUCTION AND MOTIVATION

In the spring of 1987, I made the short trip from Cornell to Syracuse University to give a seminar and talk with members of the Syracuse relativity group. I was particularly excited about the prospect of meeting Demetrios Christodoulou since we had both been studying the model problem of spherically-symmetric collapse of a minimally coupled, massless scalar field. Even though our approaches to the problem were apparently completely different—his purely analytical [1], mine purely numerical [2]—I was hoping that he would have ideas and insights which would suggest new directions for my research. I was not disappointed.

Christodoulou had been thinking about the model in terms of one-parameter families of spacetimes generated from the evolution of initially-imploding packets of scalar field. For each family, the single parameter, p , was to characterize the “strength”

of the initial data and hence control the degree of non-linearity in the evolution. The idea was that a generic family would “interpolate” between a weak-field limit, $p \rightarrow p_{\text{weak}}$, characterized by linear propagation of the scalar field on an essentially flat background, and a strong-field regime, $p \rightarrow p_{\text{strong}}$, where the gravitational self-interaction of the scalar field was sufficient to form a black hole. For each family, in varying p from p_{weak} to p_{strong} , there would be a critical parameter value, p^* , where a spacetime containing a black hole first appeared. The question Christodoulou posed was whether, in the generic case, this smallest black hole would have *finite* or *infinitesimal* mass.

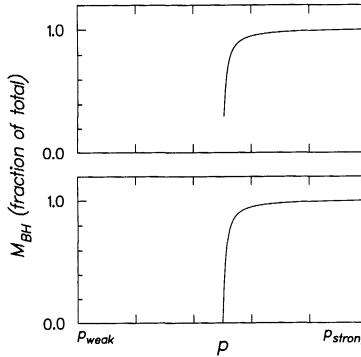


Figure 1. Schematic representation of the question posed by Christodoulou: for a generic interpolating 1-parameter family of solutions of the coupled Einstein/scalar equations, would the smallest black hole mass be *finite* (top) or *infinitesimal* (bottom)?

At the time, Christodoulou had suspicions that the smallest black hole in a generic family would have finite mass. I was intrigued by the possibility of investigating this question numerically, yet not very optimistic about obtaining a definitive numerical answer. On the one hand, were Christodoulou’s suspicions correct, then it seemed that a numerical study of various one-parameter families could do no better than provide corroborating evidence on a case-by-case basis. Conversely, if infinitesimal black holes were the norm, how could a numerical code be used to conclusively establish this fact, given that any specific calculation must necessarily be carried out at a finite resolution?

Despite my initial pessimism, I concentrated my subsequent numerical efforts on exploring the nature of critical points generated by parameterized collapse. Eventually, much to my surprise, a remarkable picture of the transition to and from black hole spacetimes emerged. I found that the behaviour of the model in the near-critical regime was characterized by a number of non-linear phenomena, including exponential sensitivity to initial conditions, a type of discrete scale invariance, universal power-law dependence of black hole mass on parameter space displacement $|p - p^*|$, and, in answer to Christodoulou’s question, *infinitesimal* black hole mass at a critical point.

This paper is a summary of both the underlying methodology and the results of my study of these critical phenomena. Following a brief review of the formalism and equations of motion for the model in the next section, I discuss my basic numerical approach and the Berger and Oliger mesh-refinement algorithm (which was instrumental in obtaining the results) in Section 3. The critical phenomena *per se* are described in Section 4 and Section 5 contains a discussion of the analysis of error in the numerical calculations. Related work is sketched in Section 6 which is followed by a few concluding remarks.

2. FORMALISM AND EQUATIONS OF MOTION

In this section I briefly review the formalism and equations of motion I have used in studying spherically-symmetric scalar field collapse [2, 3, 4, 5]. Except for brief comments in Section 6, the remainder of this paper concerns the case where the field ϕ , is real, massless, and minimally coupled to the gravitational field.

2.1. Flat spacetime: the weak-field limit

Adopting the usual polar-spherical coordinates, (t, r, θ, φ) , and employing geometric units, $G = c = 1$, the Minkowski line element is

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2) \equiv -dt^2 + dr^2 + r^2 d\Omega^2. \quad (1)$$

In these coordinates, the equation of motion for a massless scalar field, ϕ , may be written as the usual one-dimensional wave equation for the quantity $r\phi$:

$$\square\phi = 0 \implies \frac{\partial^2}{\partial t^2}(r\phi) = \frac{\partial^2}{\partial r^2}(r\phi). \quad (2)$$

Clearly, at all times, the general solution of (2) can be written as a sum of ingoing (imploding) and outgoing (exploding) contributions:

$$r\phi(r, t) \sim u(r+t) + v(r-t). \quad (3)$$

If the initial data is specified in terms of an initially ingoing profile, $f(r)$, and an initially outgoing profile, $g(r)$, so that

$$r\phi(r, 0) = f(r) + g(r) \quad (4)$$

$$r\phi(r, 0) = \frac{df}{dr}(r) - \frac{dg}{dr}(r) \quad (5)$$

then it is straightforward to show that the complete solution of (2) is given by

$$\begin{aligned} r\phi(r, t) &= f(r+t) + g(r-t) & r \geq t \\ &= f(t+r) - f(t-r) & r < t. \end{aligned} \quad (6)$$

Perhaps the most notable feature of the dynamics in the linear regime is that the solutions of (2) are non-dispersive; any and all “structure” encoded in the initial data, $r\phi(r, 0)$ is propagated without change to (future null) infinity.

2.2. Curved spacetime

With an appropriate choice of r and t coordinates [2, 3, 4, 5], the metric for an asymptotically-flat, spherically-symmetric, time-dependent spacetime may be written as

$$ds^2 = -\alpha^2(r, t) dt^2 + a^2(r, t) dr^2 + r^2 d\Omega^2. \quad (7)$$

Equivalently we may write

$$ds^2 = -\alpha^2(r, t) dt^2 + \left(1 - \frac{2m(r, t)}{r}\right)^{-1} dr^2 + r^2 d\Omega^2. \quad (8)$$

This second form, written as it is in terms of the *mass aspect*, $m(r, t)$, emphasizes that this coordinate system can be viewed as a natural generalization of Schwarzschild coordinates to time-dependent geometries. Note that the radial coordinate, r , measures proper surface-area and is thus covariantly defined. On the other hand, the time coordinate, t , has no particularly relevant geometrical interpretation, except as $r \rightarrow \infty$, where it measures proper time. However, for initially regular spacetimes (such as those described below), there is a geometrically preferred time variable in spherical symmetry, namely the proper time, T_0 , of an observer fixed at $r = 0$:

$$T_0 \equiv \int_0^t \alpha(0, \tilde{t}) d\tilde{t}. \quad (9)$$

The fact that the critical phenomena described below were discovered via computations in r and t , but are most naturally described in terms of r and T_0 is a particularly pleasing manifestation of general covariance.

In order to compute the dynamics of the scalar field, I find it convenient to introduce auxiliary variables, Φ and Π :

$$\Phi(r, t) \equiv \frac{\partial \phi}{\partial r}(r, t), \quad \Pi(r, t) \equiv \frac{a}{\alpha} \frac{\partial \phi}{\partial t}(r, t). \quad (10)$$

With these definitions, the following equations are sufficient to determine the evolution of the coupled scalar and gravitational fields:

$$\frac{\partial \Phi}{\partial t} = \frac{\partial}{\partial r} \left(\frac{\alpha}{a} \Pi \right), \quad (11)$$

$$\frac{\partial \Pi}{\partial t} = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\alpha}{a} \Phi \right) = 3 \frac{\partial}{\partial(r^3)} \left(r^2 \frac{\alpha}{a} \Phi \right), \quad (12)$$

$$\frac{1}{a} \frac{da}{dr} + \frac{a^2 - 1}{2r} - 2\pi r (\Pi^2 + \Phi^2) = 0, \quad (13)$$

$$\frac{1}{\alpha} \frac{d\alpha}{dr} - \frac{1}{a} \frac{da}{dr} + \frac{1 - a^2}{r} = 0. \quad (14)$$

Equations (11) are equivalent to the covariant wave equation for ϕ , (13) is the Hamiltonian constraint, used to determine the radial metric function, a , and (14) is the *polar slicing* [6] condition which fixes the *lapse function*, α , on each $t = \text{constant}$ slice.

The system of equations (11–14), and any solutions thereof, are invariant under the rescalings

$$(r, t) \rightarrow (kr, kt) \quad \text{for real } k > 0, \quad (15)$$

which reflects the absence of any intrinsic mass/length scale in the model. Since these transformations have the potential to obfuscate issues such as the minimum black hole mass in a one-parameter family, it is important to understand that the critical dynamics is naturally expressed in terms of variables which are *form-invariant* with respect to these rescalings. Suitable form-invariant variables for the scalar field are given by

$$X(r,t) \equiv \sqrt{2\pi} \frac{r}{a} \Phi = \sqrt{2\pi} \frac{r}{a} \frac{\partial \phi}{\partial r}, \quad (16)$$

$$Y(r,t) \equiv \sqrt{2\pi} \frac{r}{a} \Pi = \sqrt{2\pi} \frac{r}{a} \frac{\partial \phi}{\partial t}, \quad (17)$$

as well as by the field, ϕ , itself. The geometric quantities $a(r,t)$, $m(r,t)/r$ and $dm(r,t)/dr$ (among others) are also form-invariant. The last of these has a particularly simple expression in terms of the form-invariant wave amplitudes, X and Y :

$$\frac{dm}{dr} = X^2 + Y^2. \quad (18)$$

Thus the total mass of the spacetime, M_{ADM} , may be computed using

$$M_{\text{ADM}} = \int_0^\infty \frac{dm}{dr} dr = \int_0^\infty X^2 + Y^2 dr. \quad (19)$$

In some respects, the (r,t) coordinates I have adopted might seem to be a poor choice for studying gravitational collapse, particularly since it is well known that such coordinates can not penetrate event horizons [6]. However, there are significant computational advantages to using this system (notably the simplicity of the equations of motion, (11–14)) and, more importantly, there is no significant operational disadvantage for studying the *formation* of black holes. In fact, black hole formation is very easily detected in any specific calculation by monitoring, for example, the quantity $2m/r$. For black hole spacetimes, this function will rapidly asymptote to the value 1 for some specific $r = R_{\text{BH}}$, from which the mass of the hole $2M_{\text{BH}} = 2R_{\text{BH}}$, can be immediately deduced.

3. NUMERICAL TECHNIQUES

3.1. Finite differencing: basic approach

My numerical analysis of equations (11–14) is based on straightforward, second-order finite difference techniques [2, 3]. Fig. 2 shows a schematic depiction of a typical finite difference grid which might be used to generate an approximate solution of the model problem described above. The grid is *uniform*, which means that both the radial and temporal mesh spacings (Δr and Δt) are constant throughout the computational domain. Furthermore, for hyperbolic problems, such as the coupled Einstein/scalar system, it is natural to use difference grids which are characterized by a *single* discrete scale, h . Thus, I set $\Delta r = h$ and then define $\Delta t \equiv \lambda \Delta r = \lambda h$ where the Courant parameter, λ , is understood to be fixed whenever h is varied.

I use a standard finite-difference notation for functions defined on the mesh

$$f_j^n \equiv f(r_j, t^n) = f(j \Delta r, n \Delta t) = f(jh, n\lambda h), \quad (20)$$

and note that, particularly for issues such as convergence, stability and accuracy, it is useful to adopt the point of view that grid functions are actually defined everywhere (via appropriately high-order interpolation in the mesh, for example) on that part of the continuum covered by the finite-difference mesh, rather than only at mesh points [2]. Consistent with this interpretation, I will sometimes use a “ $\hat{\cdot}$ ” notation to denote a grid function, $\hat{f} \equiv \hat{f}(r, t; h)$, which approximates some continuum function, $f \equiv f(r, t)$. It is also important to keep in mind that when performing finite-difference calculations we are, after all, most interested in the continuum limit, $h \rightarrow 0$. Even though bounds on our computational resources will always restrict (sometimes severely so) the extent to which we can approach this limit, we nonetheless always have at our disposal the extremely simple, general and powerful tool of *convergence testing*. Namely, by performing the same computation at different resolutions, h , we can *always* make some statement about how close we are to the continuum limit (i.e. how large our errors are) and evidence of

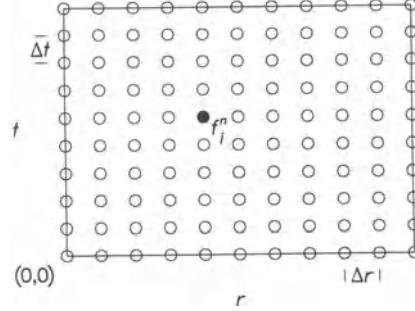


Figure 2. Basic structure of a typical finite-difference mesh used in the computations described in the text. Note that both the radial and temporal mesh spacings, Δr and Δt are constant.

this sort of analysis is a key feature anyone (numericist or not) should look for when attempting to make a judgement concerning the reliability of some new finite-difference result.

I use the following second-order ($O(h^2)$) discretization of (11-14):

$$\frac{\Phi_j^{n+1} - \Phi_j^{n-1}}{2 \Delta t} = \frac{1}{2 \Delta r} \left[\frac{\alpha_{j+1}^n}{a_{j+1}^n} \Pi_{j+1}^n - \frac{\alpha_{j-1}^n}{a_{j-1}^n} \Pi_{j-1}^n \right] \quad (21)$$

$$\frac{\Pi_j^{n+1} - \Pi_j^{n-1}}{2 \Delta t} = \frac{3}{r_{j+1}^3 - r_{j-1}^3} \left[r_{j+1}^2 \frac{\alpha_{j+1}^n}{a_{j+1}^n} \Phi_{j+1}^n - r_{j-1}^2 \frac{\alpha_{j-1}^n}{a_{j-1}^n} \Phi_{j-1}^n \right] \quad (22)$$

$$\begin{aligned} \frac{A_{j+1}^{n+1} - A_j^{n+1}}{\Delta r} &+ \frac{\exp(A_{j+1}^{n+1} + A_j^{n+1})}{2r_{j+\frac{1}{2}}} - \frac{1}{2r_{j+\frac{1}{2}}} \\ &- 2\pi r_{j+\frac{1}{2}} \left[\left(\frac{1}{2} (\Pi_{j+1}^{n+1} + \Pi_j^{n+1}) \right)^2 + \left(\frac{1}{2} (\Phi_{j+1}^{n+1} + \Phi_j^{n+1}) \right)^2 \right] = 0 \end{aligned} \quad (23)$$

$$\frac{\alpha_{j+1}^{n+1} - \alpha_j^{n+1}}{\Delta r} + \frac{1}{2} (\alpha_{j+1}^{n+1} + \alpha_j^{n+1}) \left[\frac{1 - \left(\frac{1}{2} (\alpha_{j+1}^{n+1} + \alpha_j^{n+1}) \right)^2}{r} - \frac{2 (\alpha_{j+1}^{n+1} - \alpha_j^{n+1})}{\Delta r (\alpha_{j+1}^{n+1} + \alpha_j^{n+1})} \right] = 0 \quad (24)$$

where

$$A_j^n \equiv \ln a_j^n. \quad (25)$$

Note that in this scheme, all of the grid functions, Φ_j^n , Π_j^n , a_j^n , A_j^n and α_j^n are defined on the same finite-difference mesh and, consequently, at all mesh locations. This is in contrast to the strategy often adopted in numerical relativity work where one introduces a collection of meshes which are “staggered” (offset in space and or time, typically by one half-mesh-spacing) and then defines any specific grid function on whichever component mesh will best facilitate the “centering” of the equation of motion for that function. I have purposefully *not* adopted this approach in the current work since it would significantly complicate the incorporation of the resulting difference scheme into the mesh-refinement algorithm described in Section 3.3.

A solution of the finite-difference system (21–25) begins with the specification of effectively arbitrary values, Φ_j^0 and Π_j^0 , from which consistent values of the geometric variables a_j^0 and α_0^j are determined using (24) and (25). Advanced values Φ_j^1 , Π_j^1 , a_j^1 and α_j^1 are then computed via an iterative solution of the set of equations derived from (21–25) by making the substitutions $n-1 \rightarrow n$, $n \rightarrow n+1/2$ and $\Delta t \rightarrow \Delta t/2$, where, for any function f ,

$$f_j^{n+\frac{1}{2}} \equiv \frac{1}{2} (f_j^n + f_j^{n+1}). \quad (26)$$

Once the initial data have been set, the system may be (repeatedly) advanced from time t^n to t^{n+1} by (i) computing advanced values for the scalar field variables Φ_j^{n+1} and Π_j^{n+1} from (21) and (22), (ii) solving the discrete Hamiltonian constraint (23) for a_j^{n+1} using a pointwise Newton iteration [2] to handle the nonlinearity, and (iii) solving the slicing condition (24) for the lapse values α_j^{n+1} .

3.2. Finite differencing: Richardson expansions

As I have already remarked, when solving differential equations using finite-difference techniques, the issue of convergence is a key concern. Many years ago, Richardson [7] made a crucial observation concerning the convergence behaviour of finite-difference approximations which I review here due to its fundamental importance and utility for finite-difference work in general, and for the current work in particular [2].

Consider a differential equation

$$L u(r, t) = 0, \quad (27)$$

where L is some differential operator, and let

$$\hat{L}(h) \hat{u}(r, t; h) = 0 \quad (28)$$

be a finite-difference approximation of (27) which, like the system (21–25), is characterized by a single discretization scale, h . Assume that the finite-difference approximation

is $O(h^2)$ and involves only centred-difference operations. By definition this means that the truncation error, $\tau(r, t; h)$, obtained by applying the finite-difference operator, \hat{L} , to the continuum solution, u , may be written as

$$\tau(r, t; h) \equiv \hat{L}(h) u(r, t) = h^2 \tau_2(r, t) + h^4 \tau_4(r, t) + \dots \quad (29)$$

Here, $\tau_2(r, t), \tau_4(r, t) \dots$ are h -independent function with the same smoothness as derivatives of u of appropriately high order. What Richardson observed was that under quite general conditions, we can expect the difference solution, $\hat{u}(r, t; h)$, to admit an asymptotic ($h \rightarrow 0$) expansion of the form

$$\hat{u}(r, t; h) = u(r, t) + h^2 e_2(r, t) + h^4 e_4(r, t) + \dots \quad (30)$$

The crucial feature of such a *Richardson expansion*, is that the error functions $e_2(r, t)$, $e_4(r, t)$ are also h -independent functions with smoothness comparable to some derivative of the continuum solution, u . Among other things, this implies that by performing the same computation at different resolutions h , the error functions, e_2, e_4, \dots become no less computable than the solution, u , itself. This in turn suggests a useful (if often heuristic) way of viewing finite difference calculations—when we solve a finite difference equation such as (28), we are in effect solving an infinite system (hierarchy) of differential equations for all of the error terms as well as for the fundamental solution.

The extent to which expansions of the form (30) provide an accurate representation of grid functions in specific calculations depends on a multitude of factors, and although *a priori* analytic investigations of this question are certainly possible for any given differential system and difference approximation, the notion of Richardson-expandability can also be used profitably in an *a posteriori* mode. Briefly, this means that we follow a few basic principles in constructing our difference schemes and then assume that our numerical results will be expandable. Empirical verification of this assumption then provides us with strong evidence that we are producing approximate solutions of the differential system, as well as with error estimates for specific calculations. Finally, even if we “perturb” the structure of our computations so much (by introducing mesh-refinement, for example) that *global* expansions like (30) no longer appear to exist, this will almost certainly be due to the fact that the perturbations will have effectively introduced “non-smooth” terms into the hierarchy of error equations. In these circumstances it will often still be sensible to assume that grid functions are at least *locally* expandable throughout most of the computational domain. It is precisely this assumption which is used in the Berger and Oliger algorithm which I now summarize.

3.3. The Berger and Oliger mesh-refinement algorithm

If computers were infinitely fast and equipped with unbounded memory, it would be a trivial matter (modulo stability) to finite-difference any differential system on a uniform mesh with h chosen sufficiently small to yield a solution with error below any *a priori* specified tolerance. However, in the real world, where computational resources are finite, the choice of discretization scale for most systems of partial differential equations (PDEs) is still very much constrained, and the issue of how to “distribute the resolution where it is needed” is of paramount importance.

In 1984, M. Berger and J. Oliger [8] proposed a general mesh-refinement algorithm for hyperbolic PDEs wherein the mesh scale becomes a local quantity which changes in response to the development of solution features. Mesh-refinement in the Berger

and Oliger (BO) algorithm is implemented using a nested, hierarchical approach very reminiscent of the multi-level adaptive techniques (MLAT) long advocated by Brandt [9] for the near-optimal solution of PDEs and other continuum problems. Specifically, the BO algorithm employs a discrete sequence of mesh scales, h_l , $l = 0 \dots L$ which satisfy

$$h_0 = \rho_0 h_1 = \rho_1 h_2 = \dots = \rho_{L-1} h_L. \quad (31)$$

Here, h_0 is the coarsest (base) scale, h_L is the finest scale, and the ρ_l are integer-valued refinement ratios which typically satisfy $\rho_0 = \rho_1 = \dots = \rho_{L-1} = \rho$ for some fixed refinement ratio, ρ . At any fixed integration time, the (spatial) computational domain is covered by a set of grids

$$\left\{ G_l^i \right\} \quad l = 0 \dots L \quad i = 0 \dots \text{ngrid}_l \quad (32)$$

where the subscript l again labels the level of discretization while the superscript i

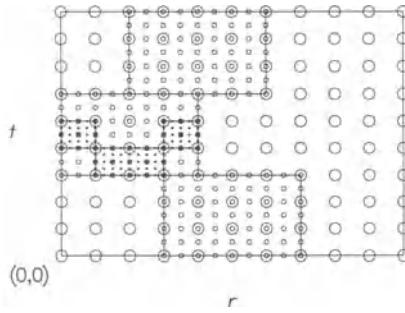


Figure 3. Schematic representation of a typical grid structure admitted by the Berger and Oliger mesh-refinement algorithm. In this example, there are three distinct levels of $2 : 1$ refinement ($h_0 = 2h_1 = 4h_2$): grid points at finer levels are marked with smaller symbols. In general the spatial distribution of grids is dynamically (adaptively) determined as the computation proceeds in time.

labels the $\text{ngrid}_l \geq 1$ disjoint level- l finite-difference grids extant at that time. Each component grid is itself uniform in the sense defined in Section 3.1, and each grid G_l^i , $l > 0$ must be properly contained within a parental grid $G_{l'}^{i'}$, $l' < l$ and (in my implementation) aligned with this parent such that in the region of mutual coverage, grid points of the parent coincide with grid points of the child. Fig. 3 is a schematic illustration of the type of grid structure admitted by the BO algorithm for the case $L = 2$, $\rho = 2$. Note that refinement is performed in time as well as in space, and that the grid structure (spatial distribution of grid points) is a dynamic quantity.

Refinements (grids G_l^i with $l > 0$) are introduced and removed in the BO algorithm in an attempt to keep the local truncation error of the discrete equations below some *a priori* specified threshold. Periodically (every few time steps), and on every grid G_l^i , $l < L$, an estimate of the local truncation error is generated on that grid and then,

if necessary, new grids G_{l+1}' are introduced to efficiently cover those regions where the estimated error exceeds the threshold. Truncation error estimation is based on the principle of Richardson expandability discussed in the previous section. For the sake of exposition, assume that we are integrating a single time-dependent differential equation, and that the difference scheme we have adopted can be written in the form

$$\hat{u}^{n+1} = \hat{Q}(h) \hat{u}^n, \quad (33)$$

where $\hat{Q}(h)$ is the one-step “update operator” at mesh scale h . Then the (one-step) local truncation error, τ^n , associated with $\hat{Q}(h)$ is defined by

$$\tau^n \equiv u^{n+1} - \hat{Q}(h) u^n. \quad (34)$$

Assuming that u admits a Richardson expansion of the form (30), which can be equivalently written as

$$u = \hat{u} + h^2 \hat{e}_2 + \dots, \quad (35)$$

it is easy to show that τ^n may be approximated as

$$\tau^n \sim \text{constant } (\hat{Q}(h) \hat{Q}(h) - \hat{Q}(2h)) \hat{u}^n. \quad (36)$$

Operationally, this means we advance the solution on G_l' two time-steps using the level l difference equations, then subtract the result of advancing a “coarsened” version of \hat{u}^n (consisting of every second mesh value) one coarse time-step using the same difference equations formulated on level $l - 1$. This technique is essentially identical to that employed, for example, in many automatic ordinary differential equation solvers. It has the great virtue of requiring only the same information which is already needed to update the difference solution on any grid and is therefore generally applicable.

In the context of the critical phenomena discussed in the next section, implementation of the BO algorithm for spherically-symmetric scalar was crucial for the automatic and efficient resolution of solution features on a wide range of spatio-temporal scales. It is, in fact, highly unlikely that I would have discovered the critical behaviour without the mesh-refinement code. As an illustrative example of the gains in computational efficiency provided by the algorithm, a typical near-critical configuration of the sort described below is computed using a base grid G_0^0 having approximately 600 points in the spatial (radial) direction, and a maximum of 7 (sometimes more) additional levels of $5 : 1$ refinement ($L = 7$, $\rho = 5$). Thus, a uniform fine grid at the finest available level of resolution, h_7 , would have about 5×10^7 points and a typical evolution would require roughly the same order of time steps for a total of perhaps 10^{15} events. In practice, even though the finest level of discretization is used, the total number of radial grid points never exceeds 2500 and something of the order of 10^7 events are computed. The basic factor of 10^8 or more provided by the mesh-refinement algorithm completely dwarfs other performance improvements achieved through vectorization and other traditional forms of optimization, as well as any gains provided by the continuing growth in available computer resources during the time the research was carried out.

4. CRITICAL PHENOMENA

4.1. Parameter space surveys

As intimated in Section 1, the discovery of critical behaviour [4, 5] in the Einstein/scalar model was a direct result of an attempt to determine the nature of the black hole

transition point (critical point) for generic one-parameter “interpolating” families of solutions of equations (11–14). Typically, the solutions I have studied are generated by evolving parameterized, initially-ingoing ($g(r) = 0$ in (4)) single pulses (shells) of scalar field. A canonical example is a “generalized Gaussian” profile:

$$\phi(r, 0; \phi_0, r_0, \delta, q) = \phi_0 r^3 \exp(-((r - r_0)/\delta)^q). \quad (37)$$

Appropriate one parameter families may be generated from this profile by varying any of ϕ_0, r_0, δ or q while keeping the other values fixed. We can naturally associate two distinct length scales with any pulse of this type. The first is simply the physical thickness, L , of the pulse (a few δ for the above profile), the second is the Schwarzschild radius, $R_S \equiv 2M_{\text{ADM}}$ (see equation (19)), of the configuration. When $L \gg R_S$ ($\phi_0 \rightarrow 0$ for example), we are in the weak-field regime—the scalar and gravitational fields decouple and the dynamics is described by flat spacetime solutions of the spherically-symmetric wave equation (Section 2.1). In this case, an ingoing pulse like (37) simply implodes through $r = 0$, then escapes to infinity in a completely non-dispersive and non-dissipative manner. However, for $L \approx R_S$, strong-field (non-linear, self-gravitating) effects become significant (particularly during the period of maximal implosion) and black hole formation becomes a possibility.

Given any 1-parameter family of initial data, $\phi(r, 0; p)$, which is known to completely disperse for $p \leq p_{\text{LO}}$ and form a black hole for $p \geq p_{\text{HI}}$, it is a straightforward matter to use a bisection technique to numerically determine the critical parameter value, p^* ($p_{\text{LO}} < p^* < p_{\text{HI}}$), which represents the threshold of black hole formation for the family.

4.2. Sensitivity to initial conditions

Shortly after implementing the adaptive-mesh algorithm for the scalar model, I undertook a rough search for a critical point using initial data $\phi(r, 0; \phi_0)$ of the form (37). Due to my prior discussion with Christodoulou, I was trying to make very small black holes, and although those early runs at Cornell—on an IBM 3090 running CMS—were nowhere near as systematic as what I would eventually carry out, there were already indications that the resolution available to the algorithm, rather than the details of the initial data, determined the minimum black hole mass. I subsequently moved to CITA in Toronto and began using Silicon Graphics (SGI) machines which not only ran a different operating system (UNIX), but also employed a type of 8-byte floating-point arithmetic different from the 3090’s. When I ran the code on the SGI machine using the initial data specification which had produced the smallest black hole in the runs made on the IBM architecture, I discovered to my horror that no black hole formed! Initially I suspected that there was a flaw in my implementation of the BO algorithm, particularly since the grid structures produced by the two runs looked completely different at late times. This suspicion was compounded by the generic appearance of “glitches” in the late-time waveforms from near-critical computations. What bothered me most, however, was how difficult it was to analyze what was going on near the critical point. It was clear that there was structure developing on small spatial scales but it formed on correspondingly short time scales, and the primitive state of my graphics (visualization) capabilities made getting a clear picture of the dynamics a slow and tedious process.

This prompted me to invest considerable effort in the development of a “visualizer” specifically designed to display and analyze time series of one dimensional data. A chief

feature of this tool was that it was implemented using a *client/server* approach wherein various clients (such as the adaptive-mesh code running on some machine) sent data directly over the network to the visualizing server which ran on a graphics workstation and presented a uniform, interactive interface for the display and manipulation of the data. The server readily handled multiple data-streams so that I could easily monitor, for example, the dynamics of a particular grid function computed by the adaptive-mesh algorithm on distinct discretization levels. With this tool I quickly tracked down and cured the “glitching” problem (which turned out to be caused by a rather subtle and mostly benign instability), but more importantly, I was able to take a closer look at the details of the dynamics near the critical point.

This closer examination soon bolstered my confidence in the results produced by the mesh-refinement algorithm near criticality. In accord with the early work at Cornell, I found that by letting the algorithm use finer and finer resolution, I could observe phenomena—including black hole formation—on smaller and smaller spatio-temporal scales. However, this was true only up to a point. Provided the maximum permitted resolution, h_L , was fine enough, I found that a bisection search for a critical value, p^* , could be carried out to the limit of machine precision. Thus, the final values of p_{LO} and p_{HI} produced by the search would differ fractionally by perhaps a part in 10^{15} , both computations would appear well resolved, yet one would form a black hole (typically with a mass of about $10^{-6} M_{ADM}$), while the other would not. After investigating critical points for several separate families and always observing the same behaviour, it dawned on me that logarithmic transformations would simplify the description of the phenomenology of the model near the critical point. Specifically, I observed that if I re-expressed the initial data in terms of a new parameter π defined by

$$\pi \equiv \ln |p - p^*|, \quad (38)$$

then the amount of small-scale structure which developed was basically linear in π .

Linear sensitivity to π clearly implied exponential sensitivity to initial conditions in the model as $p \rightarrow p^*$, and Figs. 4–6 show typical evidence of this sensitivity. Each frame in these plots displays the late-time (some fixed final time, t_f) profile of the quantity $2m/r$ from a separate computation using initial data of the form (37). In Fig. 4, where each frame shows $2m/r$ plotted versus the radial coordinate r , the development of structure in the trailing (left-most) edge of the profile is apparent for the first few subcritical ($p < p^*$) computations, but then becomes impossible to resolve on the scale of the graphic. In $\ln(r)$ coordinates, however (Fig. 5), the regular development of features as a function of $\ln |p - p^*|$ is vividly apparent. A similar development can also be seen in the supercritical ($p > p^*$) regime where black holes form (Fig. 6).

Interestingly enough, the presence of exponential sensitivity in the model provides an explanation (albeit heuristic) for the “machine-dependence” of near-critical calculations. The mesh-refinement algorithm adds a significant degree of complexity and effective non-linearity to the numerical computations, particularly in the near-critical regime where many levels of discretization are employed and regridding operations are extensive. Refinements are introduced when truncation error estimates exceed a certain threshold and it thus becomes quite probable that on one machine, some refinement ends up with one more (or less) grid point than its counterpart on another architecture, due to the different floating-point arithmetic. Given the nature of the mesh-refinement algorithm, the effect of adding a single extra grid point amounts to a huge amplification of this roundoff-level deviation and, since the problem is so sensitive, the subsequent

structure of the two computations, especially on scales smaller than that where the discrepancy started, can be quite different (see Section 5.1). Nevertheless, for any given combination of initial data family $\phi(r, t; p)$ and floating-point hardware, I have always found an absolutely smooth and continuous development of small-scale structure as $p \rightarrow p^*$. Thus, at least in the sense of exhibiting exponential divergence of initially-close trajectories, the model does *not* seem to exhibit chaos as a critical point is approached.

4.3. Scale-periodicity (echoing)

A striking aspect of the plots in Figs. 5 and 6 is the extremely *regular* nature of the

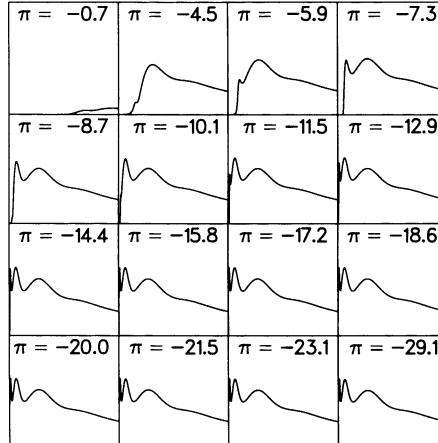


Figure 4. Late-time profiles of $2M/r$ versus r from a sequence of 16 subcritical computations with family-parameter, ϕ_0 , logarithmically “zooming-in” to the critical value ϕ_0^* . Each profile is labeled with the corresponding logarithmically rescaled parameter value, $\pi \equiv \ln |\phi_0 - \phi_0^*| + \kappa$. In each frame, the range of $2m/r$ is 0.00–0.26, and in all cases the displayed profiles are almost entirely outgoing. The formation of additional structure at the trailing edge (left side of each frame) due to nonlinear interaction between the scalar and gravitational field is difficult to observe after the first few frames due to the use of the “bare” radial coordinate, r , as the abscissa.

structure which appears in the profiles when a logarithmic radial coordinate is used as the independent variable. This structure is a manifestation of the phenomenon of *scale-periodicity*,¹ which is perhaps the most intriguing feature of the critical behaviour in the model. Specifically, all of the empirical evidence I have found concerning near-critical evolution is consistent with the following conjecture [4, 5]: the strong-field, precisely-critical behaviour of the model is described by a unique (up to trivial rescalings $r \rightarrow kr$, $t \rightarrow kt$) critical dynamics which is invariant under a certain *discrete scaling*

¹I am grateful to D. Christodoulou for suggesting this terminology.

symmetry. For example, if I describe the scalar field dynamics using the variables X and Y defined by (16-17), and express these variables as functions of the *logarithmic, physical* independent variables ρ and τ :

$$\rho \sim \ln(\text{proper radius}) \equiv \ln(r) + \kappa, \quad (39)$$

$$\tau \sim \ln(\text{proper central time}) \equiv \ln(T_0^* - T_0) + \kappa, \quad (40)$$

where T_0^* will be defined shortly, then the conjecture is that there are unique sequences of critical configurations, $X^*(\rho, \tau)$, $Y^*(\rho, \tau)$, which satisfy

$$X^*(\rho \pm \Delta, \tau \pm \Delta) \simeq X^*(\rho, \tau), \quad (41)$$

$$Y^*(\rho \pm \Delta, \tau \pm \Delta) \simeq Y^*(\rho, \tau). \quad (42)$$

Here, Δ is an apparently universal (family-independent) constant with an empirically

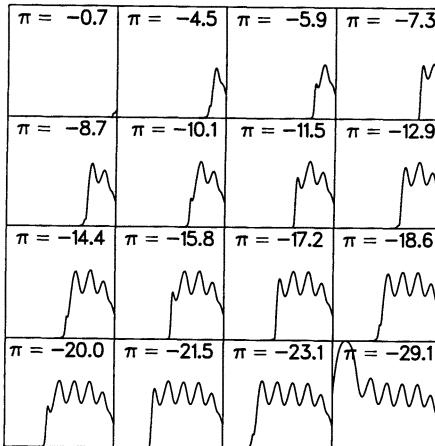


Figure 5. The same data shown in Fig. 4 is displayed here, but each profile is now plotted using a logarithmic radial coordinate. The basically linear development of additional structure in the $2m/r$ profile as a function of π —indicative of exponential sensitivity to initial conditions near the critical point—is evident.

determined value, $\Delta \approx 3.4$. Scale-periodicity is similarly conjectured for all other form-invariant (see Section 2.2) quantities—including ϕ , $r^2\phi^\mu\phi_\mu$, dm/dr and m/r .

Physically, scale-periodicity means that for a precisely-critical configuration, an infinite series of strong field “echoes” is generated from the recurrence of strong-field evolution on ever decreasing spatio-temporal scales. Each echo is produced on a scale some $e^{3.4} \approx 30$ times smaller than that associated with its predecessor and the critical dynamics “accumulates” at some *critical central proper time*, T_0^* . Since the associated oscillations in the scalar field occur on these increasingly finer scales, both the “kinetic energy” of the scalar field, $\phi^\mu\phi_\mu$, and the scalar curvature, R , of the spacetime,

$$R = -G = -8\pi T = 8\pi\phi^\mu\phi_\mu, \quad (43)$$

get driven to infinite values at the critical event, $r = 0$, $T_0 = T_0^*$. A precisely-critical spacetime is therefore necessarily singular; the singularity itself is massless and most likely naked.

The phenomenon of scale-periodicity (combined with reasonable assumptions concerning the behaviour of any mass-energy in the scalar field which has “echoed”—namely that it almost all disperses to infinity) also provides an unambiguous mechanism for producing black holes of arbitrarily small mass in any one-parameter interpolating family. In effect, the seemingly intractable problem of constructing an infinitesimal-mass black hole via numerical methods is replaced by the quite tractable problem of demonstrating that the equations of motion for the model admit an asymptotically

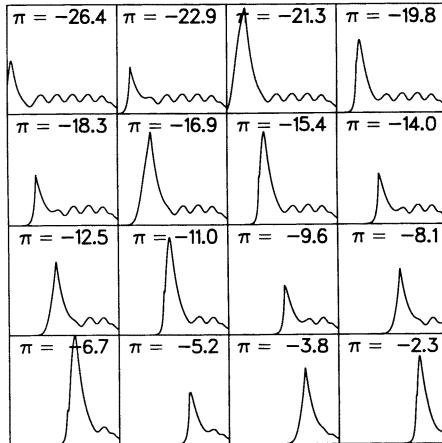


Figure 6. Late-time profiles of $2M/r$ versus $\ln(r)$ from a sequence of 16 supercritical computations with family-parameter, ϕ_0 , logarithmically “zooming-out” from the critical value ϕ_0^* . Here, $2m/r$ ranges from 0.00 to 1.00 and each generated spacetime contains a black hole whose surface is located near the cusp in the corresponding profile. Exponential sensitivity to initial conditions is again apparent.

scale-periodic solution. (In fact, a scale-periodic ansatz would probably be a good starting point for a numerical treatment of the critical solution which is more accurate than that provided by the current mesh-refinement algorithm.) Of course, without a general theorem regarding the approach to the conjectured scaling solution, scale-periodicity can only be verified on a case-by-case basis. In this regard I can only state that I have done my best to produce interpolating families which do not exhibit the scale-periodicity (41-42), but have always failed to do so [4].

4.4. Black hole mass-scaling

In the summer of 1990, having identified the critical dynamics (including the same scale-periodicity) in several distinct one-parameter families of solutions, I finally struck on the idea of examining the behaviour of the black hole mass, M_{BH} , as a function

of parameter value in supercritical calculations. Since I had by this time studied the critical solution in considerable detail, I was used to thinking about the calculations in terms of the logarithmically rescaled parameter value, $\pi \equiv |p - p^*|$, and thus it seemed only natural to set up a series of computations to compute $\ln(M_{\text{BH}})$ for a uniformly distributed set of π values. I had vague hopes that I would discover some “bizarre” behaviour, such as non-monotonicity, in $\ln(M_{\text{BH}}(\pi))$ since the behaviour of the mass aspect as a function of π seemed quite complicated (Fig. 6). Much to my astonishment, however, the results of those calculations looked like those displayed in Fig. 8: there was a strikingly *linear* relation between $\ln(M_{\text{BH}})$ and $\ln|p - p^*|$, with a coefficient of proportionality of about 0.37. That first set of computations, like the one which

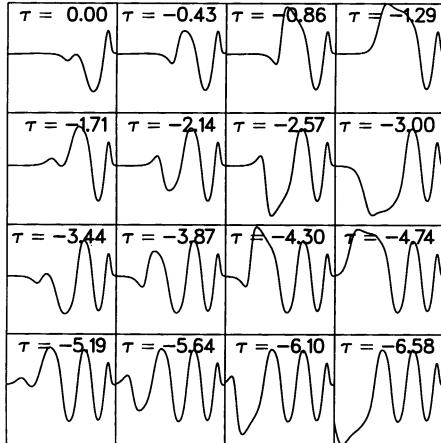


Figure 7. Near-critical evolution of X versus $\ln(\tau)$. The range of ordinate values is -0.37 – 0.33 and the temporal spacing between successive frames is uniform in logarithmic critical time, τ . The scale-periodicity conjectured in (41), with $\Delta \approx 3.4$, can be seen by comparing the trailing edges of pairs of waveforms separated by two rows in the plot. Each such pair is very nearly in the same “phase” of critical evolution, but the dynamics of the later frame (more negative τ), is occurring on a scale some $e^\Delta \approx 30$ times smaller than that of the earlier snapshot.

produced the data shown in Fig. 8, was generated by varying ϕ_0 in a family of the form (37). Due to the multi-parameter nature of this form, I was immediately able to perform independent surveys where I logarithmically varied (separately) r_0 , δ , and q away from their critical values (which were simply the fixed values I had chosen when doing the ϕ_0 computations). In all three cases, I found the same strong evidence for linearity in the $\ln(M_{\text{BH}}(\pi))$ relation, with, as far as I could tell, the same slope of about 0.37. I then performed additional studies using different 1-parameter families, and, as with the scale-periodicity in the critical solution, all the evidence suggested that the mass-scaling relationship was completely generic. Thus, these observations suggested another conjecture [4, 5]: for any interpolating family with parameter p , the black-hole

mass relation $M_{\text{BH}}(p)$ satisfied a power law as $p \rightarrow p^*$,

$$M_{\text{BH}}(p) \simeq c_f |p - p^*|^\gamma, \quad (44)$$

where c_f was a family-dependent constant, but the critical exponent, $\gamma \approx 0.37$, was universal.

Clearly, the existence of a power-law relation between black hole mass and parameter-space displacement from criticality also implied that the black-hole transition point was generically massless. What was perhaps even more intriguing was the fact that, at least for certain pulse shapes (such as those generated from the form (37)), the scaling relation (44) provided a remarkably good fit for black hole configurations

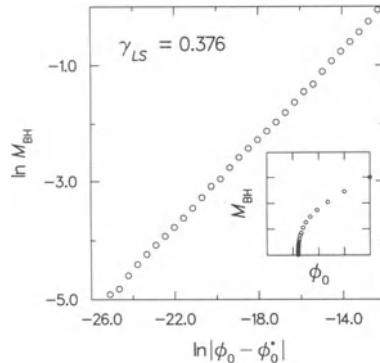


Figure 8. Typical evidence for the conjectured universal scaling relationship (44) between black hole mass and parameter-space distance from criticality. γ_{LS} is the least-squares estimate of the scaling exponent, γ , obtained via a straight-line fit to the data shown in the main graph. The inset shows the behaviour of the black hole mass as a function of (bare) parameter value (compare with Fig. 1).

well removed from the asymptotic regime, $p \rightarrow p^*$. (The most massive of the black holes depicted in Fig. 8 for example, contains nearly 90% of the total mass of the associated spacetime.)

5. ERROR ANALYSIS

The unexpected nature of the critical phenomena exhibited by the model, coupled with my use of an adaptive-mesh algorithm—which seemed like just the sort of beast capable of spontaneously generating structure on all scales—made me extremely cognizant of the need for as much analysis of the error in my numerical computations as possible. Here I will briefly discuss some of the key issues which arose in this analysis and summarize

a few of the tests which I have done at various stages in the project. All of these tests suggest that the results described above represent *bona fide* physical phenomena rather than numerical artifacts.

5.1. Sensitivity versus accuracy

One of the most important things to realize about the near-critical computations is that they are first and foremost a reflection of the sensitivity of the mesh-refinement algorithm rather than an indication of the accuracy of the results. Thus, although the algorithmic sensitivity at a computed critical point is of the order of machine precision—typically 10^{-16} – 10^{-14} —I estimate the typical fractional error in the finite-difference computations in the critical regime to be of the order 10^{-3} – 10^{-2} . This implies, for example, that the smallest black hole masses plotted in Fig. 8 are significantly less than the fluctuations in the computed value of M_{ADM} , which, analytically, is precisely conserved. How, then, can we have any faith in results concerning effects whose magnitude appears to be dwarfed by the intrinsic error in the calculations? Richardson’s observation once again provides the basis for a heuristic explanation. We argue that even near criticality, the computed solution, $\hat{\phi}^*$, should have (at least locally) a Richardson expansion

$$\hat{\phi}^* = \phi^* + h^2 e_2^* + h^4 e_4^* + \dots \quad (45)$$

where the specific form of the expansion (here assumed to be even in h) is not crucial to the argument. A generic feature of the error functions in a Richardson expansion is that they satisfy partial differential equations of very similar form and nature to the original continuum equation. Thus it is reasonable and consistent to assume that the (critical) error functions, e_2^* , e_4^* , … will manifest all of the critical attributes, such as sensitivity and scale-periodicity, displayed by $\hat{\phi}^*$ and (by assumption) ϕ^* . However the actual error, $h^2 e_2^*$ to leading order, need not be particularly small, and thus there is no fundamental problem in reconciling exquisite sensitivity with rather coarse accuracy.

5.2. Convergence tests

In my work in numerical relativity I have favoured and advocated the use of straightforward convergence tests [2, 3] whereby calculations at a series of geometrically related mesh scales $h, h/2, h/4, \dots$ are performed using the same initial data. Such tests permit the intrinsic estimation of error levels and convergence rates even in time-dependent, non-linear calculations. This approach is not trivial to implement for near-critical configurations computed by the mesh-refinement algorithm. Due to the sensitivity of the computations, convergence tests which start from the same initial profile and use the same grid structure (pattern of refinements) but different base levels of discretization, h_0 and $h_0/2$ fail miserably. Much more satisfactory results are obtained by “backwards evolution”: initial data are prescribed in terms of some previously computed critical solution and then evolved “to the past” to produce a single outgoing pulse of radiation. I have performed such tests, and although I do not observe quantitatively second order convergence, they do provide convincing evidence for convergence in the near-critical regime. I should also remark that I have demonstrated second order convergence of the mesh-refinement code in non-linear computations which are well removed from a critical point (see for example Fig. 6 of [3]).

5.3. Independent residual evaluation

Another generically applicable technique for evaluating the correctness of finite-difference results is based on the non-uniqueness of difference approximations. Thus, given a differential equation, $Lu = 0$, and a discretization, $\hat{L}\hat{u} = 0$, with an associated truncation error τ (see Section 3.2), we can introduce a distinct finite difference approximation

$$\tilde{L}(h)\hat{u} = 0, \quad (46)$$

with truncation error, τ' , given by

$$\tau'(h) \equiv \tilde{L}(h)u = h^2\tau'_2 + \dots \quad (47)$$

Assuming once again that the difference solution, \hat{u} , being tested has a (local) Richardson expansion, it is easy to see that the residual defined by

$$\tilde{L}(h)\hat{u}(h) \quad (48)$$

should be $O(h^2)$ in the limit $h \rightarrow 0$. Eschewing numerical rigour, any reasonable indication that $\tilde{L}(h)\hat{u} \rightarrow 0$ as $h \rightarrow 0$ constitutes strong evidence that the difference equation is converging to the solution of the original differential equation.

For the model problem, I have generated a second, independent discretization in an indirect fashion by discretizing the equations of motion for the variables X and Y

$$\frac{\partial X}{\partial t} = \frac{r}{a} \frac{\partial}{\partial r} \left(\frac{\alpha}{r} Y \right) - 2 \frac{\alpha a}{r} X^2 Y, \quad (49)$$

$$\frac{\partial Y}{\partial t} = \frac{1}{ra} \frac{\partial}{\partial r} (r\alpha X) - 2 \frac{\alpha a}{r} X Y^2, \quad (50)$$

which may be derived from (11–14), (16–17) and the evolution equation for the radial metric function a :

$$\frac{\partial a}{\partial t} = 2 \frac{\alpha a^2}{r} X Y. \quad (51)$$

When $O(h^2)$ approximations of (49–50) are applied to the computed values \hat{X} and \hat{Y} there is convincing evidence that the residuals tend to 0 as $h_0 \rightarrow 0$. Again, for near-critical configurations this test is best done in conjunction with “backwards” time evolution.

6. OTHER SYSTEMS

Since the initial discovery of critical behaviour in the model described above, similar phenomena have been observed in several other models of gravitational-collapse [4, 10, 11, 12]. All of these calculations have been based on the use of interpolating families of initial data: in all cases there is competition in the system (adjustable at the level of initial data) which leads to the generic appearance of critical points coincident with the threshold of black-hole formation. Here I give only a cursory summary of the current state of affairs and note that it seems likely that evidence for critical behaviour in other self-gravitating systems, as well as improved results from previously studied models, will continue to accumulate.

Critical behaviour in spherical scalar collapse has been studied for *non-minimal* coupling of the scalar radiation [4], as well as for the case where the scalar field is

explicitly self-interacting. For non-minimal coupling, all of the critical phenomena seen in the minimally coupled case are observed, with (at most) weak dependence of the critical exponents, Δ and γ , on the curvature coupling parameter. In the self-interacting case there is strong evidence that the precisely-critical picture is entirely unchanged from the massless case. In near-critical evolution, the scalar field tends to oscillate between bounded extremal values (scale-periodicity) so that “potential” terms in the wave equation, for example, tend to be bounded. At the same time the “kinetic” terms grow without bound (also due to the scale-periodicity) so that there is a natural tendency for the scalar field to become “asymptotically free” at a critical point.

Very important and interesting (not to mention computationally heroic) work has been carried out by Abrahams and Evans in the context of vacuum axisymmetric collapse using interpolating families of gravitational waves [10]. Although the reduced symmetry in their model considerably complicates many aspects of the analysis (such as determining black hole masses and identifying appropriate form-invariant quantities), the basic picture emergent from their work to date is remarkably similar to that of the spherically-symmetric scalar model. In particular, there is convincing evidence of black hole mass-scaling of the form (44) with a critical exponent $\gamma \approx 0.37$ —numerically indistinguishable from the scalar case! Moreover, there is evidence for a scale-periodic critical solution with a discrete scaling exponent, $\Delta \approx 0.6$. Thus, gravitational waves appear to echo on scales related by a factor of about 2, in contrast to roughly 30 for the scalar field, and as yet, there are no concrete proposals to explain this disparity.

Evans recently reported work he has done with Coleman using the model of a radiation fluid ($\gamma_{\text{adiabatic}} = 4/3$) in spherical symmetry [11]. This is a case where it has also been possible to make considerable analytic progress starting from an ansatz of *self-similarity* (i.e. scale-invariance rather than scale-periodicity). Power-law dependence of black hole mass on $|p - p^*|$ is again observed, with $\gamma \approx 0.36$,(!) and evidence for a unique, self-similar critical solution (consistent with the predictions of the analytic work) appears in near-critical computations.

Finally, Strominger and Thorlacius have very recently reported the discovery of universality and mass-scaling in the context of the two-dimensional, *semiclassical* RST model, which employs null matter as a matter source and has been used in simplified investigations of quantum black hole evaporation [12]. The model has the advantage of being exactly soluble and the authors have demonstrated analytically that the system exhibits universal power law mass-scaling at the critical point with $\gamma = 1/2$. They have also found a near-critical scaling solution which is interpreted to describe the formation and evaporation of an arbitrarily small black hole.

7. CONCLUDING REMARKS

The relentless advance of computational technology clearly holds the promise of revolutionizing our understanding of Einstein’s equations. A cynical reader could observe that such statements have been made for a long time now—perhaps longer than there have been numerical relativists—but there can be little argument that the rate of new information about general relativity gleaned from numerical computation has been increasing noticeably over the past few years. There can be equally little argument that we are barely beginning to scratch the surface of what is possible in numerical relativity. Thus, although numerically obtained results concerning critical phenomena in gravitational collapse may be interesting in and of themselves, they, like much of the other

work discussed in this volume, also seem significant as demonstrations of the viability of the concept of “computational laboratories” for general relativity. I hope that more researchers will be enticed to participate in the task of constructing and operating such laboratories since there is a great deal of work to be done before the true potential of numerical relativity is realized.

ACKNOWLEDGMENTS

I am grateful to J. R. Bond, D. Christodoulou, N. Kaiser, R. A. Matzner, S. A. Teukolsky, J. Thornburg, W. G. Unruh and J. Winicour for discussions and encouragement. Research was supported by the Natural Sciences and Engineering Research Council of Canada, by NSF grants PHY8806567, PHY9310083, ASC9318152 (ARPA supplemented) and by Texas Advanced Research Project grant TARP-085. All numerical computations described here were performed at the Center for High Performance Computing, University of Texas System with Cray time supported by a Cray University Research grant to Richard Matzner.

REFERENCES

- [1] Christodoulou, D., 1986, *Commun. Math. Phys.*, **105**, 337; 1986, *Commun. Math. Phys.*, **106**, 587; 1987, *Commun. Math. Phys.*, **109**, 591; 1987, *Commun. Math. Phys.*, **109**, 613.
- [2] Choptuik, M. W., 1986, *University of British Columbia PhD Thesis*, (unpublished); 1991 *Phys. Rev. D*, **44**, 3124; Choptuik, M. W., Goldwirth, D. S. and Piran, T., 1992, *Class. Quantum Grav.*, **9**, 721.
- [3] Choptuik, M. W., 1989, *Frontiers in Numerical Relativity*, (C. R. Evans, L. S. Finn and D. W. Hobill, eds. London: Cambridge University Press), p 206.
- [4] Choptuik, M. W., 1992, *Approaches to Numerical Relativity*, (R. d’Inverno ed. London: Cambridge University Press), p 202.
- [5] Choptuik, M. W., 1993 *Phys. Rev. Lett.*, **70**, 9.
- [6] Bardeen, J. M. and Piran T., 1983, *Phys. Rep.*, **96**, 205.
- [7] Richardson, L. F., 1910, *Phil. Trans. Roy. Soc.*, **210**, 307.
- [8] Berger, M. J. and Oliger J., 1984, *J. Comp. Phys.*, **53**, 484.
- [9] Brandt, A., 1977, *Math. Comput.*, **31**, 333; Brandt, A., 1982, *Lecture Notes in Mathematics, Vol 960: Multigrid Methods*, (W. Hackbusch and U. Trottenberg, eds. New York:Springer Verlag), 220.
- [10] Abrahams, A. M. and Evans, C. R., 1993 *Phys. Rev. Lett.*, **70**, 2980.
- [11] Evans, C. R., 1993 (*Talk presented at Pennsylvania State University Numerical Relativity Workshop*).
- [12] Strominger, A. and Thorlacius, L., 1993, *ITP preprint*, NSF-ITP-93-144.

COSMOLOGICAL SYSTEMS

RELATIVISTIC COSMOLOGIES

M.A.H. MacCallum¹

School of Mathematical Sciences, Queen Mary and Westfield College,
University of London, Mile End Road, London E1 4NS, U.K.

Abstract. A review of general relativistic cosmologies intended as a starting point for the more detailed discussions of the rest of the workshop is given. After a brief survey of available models, the relativistic kinematics and dynamics of cosmological models is developed. The “standard” (Friedman-Lemaître-Robertson-Walker) models are described, then the Bianchi models, and finally some inhomogeneous cosmologies.

1. AVAILABLE MODELS

For the purposes of this article ‘relativistic’ will mean ‘general relativistic’. Within this setting, there is a hierarchy of possible models which one can consider, with increasing symmetry, as shown in Table 1. Not all possibilities are shown there, only those commonly discussed. For a more complete survey of possible solutions to Einstein’s equations, see Kramer *et al.* (1980). I reviewed some of the issues I touch on in here more extensively in 1979 (MacCallum, 1979a). Other possibly useful reviews of all or a sub-part of the material covered here were given by Carmeli *et al.* (1981), MacCallum (1984), Krasinski (1990), Ferrari (1990), Griffiths (1991), and Verdaguer (1985; 1993). In the present article I focus on definitions of the different classes and the derivations of dynamical systems for further study, rather than re-considering the physical significance which I have recently reviewed elsewhere (MacCallum, 1993a, b).

A particularly important way of adding some kind of symmetry is to add a homothetic motion to the isometries. An isometry is a symmetry which preserves the metric, while in a homothety the metric is scaled by a factor; in a one-parameter group

¹E-mail address: M.A.H.MacCallum@qmw.ac.uk

of homotheties this factor can be written as an exponential e^w in terms of the group parameter w . Groups of isometries are usually considered to be generated by vector fields called Killing vectors: see e.g. Kramer *et al.* (1980) for a more complete introduction.

The evidence against the steady state models, and indeed any other model with a time-like Killing vector, comes from the observations of the magnitude-redshift relation (the Hubble law), which suggests there is expansion, the radio source counts, which suggest there has been evolution, and the existence of the cosmic microwave background radiation (CMBR) which suggests the existence of a hot dense epoch in the universe's history. While each of these arguments has been criticized, both on grounds of accuracy of the reported data and of its theoretical interpretation, almost all cosmologists are convinced that models with timelike Killing vectors are unacceptable.

The usual assumption is that on the large scale, the universe is isotropic and spatially homogeneous. This would imply that there is spherical symmetry at every point, and the geometry is then described by the Robertson-Walker (RW) metric. The name arises because Robertson and Walker independently gave derivations of the metric from the assumption of isotropy at every point: the longer names given to this class of models, attaching the names of Friedman and Lemaître (FLRW), are used to refer to general relativistic models with this geometry and the dynamics given by Einstein's field equations. The matter content must take the form of a perfect fluid (although this can be mimicked by a combination of energy-momentum tensors of other forms, as has been emphasized by Coley and Tupper (1983)) and with the usually-assumed equations of state the models have a big-bang singularity (unless there is a significant cosmological constant).

Table 1. Symmetry classes of cosmological models

Symmetry type of model	Governing equations	name
General	10 PDEs in 4 independent variables	
Spacelike Abelian two-parameter group	PDEs in 2 variables	" G_2 cosmologies"
Isotropic at 1 point	PDEs in 2 variables	Spherically symmetric
Spatially homogeneous	ODEs	Bianchi and Kantowski-Sachs models
Spatially homogeneous and isotropic	ODEs	Robertson-Walker
Homogeneous	Only algebraic equations	steady state

These models are clearly not exactly correct, since the density of matter is not uniform, and the precise sense in which they are supposed to be appropriate remains somewhat unclear (Ellis and Stoeger, 1987). From the observations, support for (near) isotropy comes from the distributions of galaxies and radio sources, and the isotropy of

the Hubble law and the cosmic background radiations. Among these last, the microwave background has the most precisely-determined isotropy: only very recently have the small residual anisotropies been observed (Smoot *et al.*, 1992). As with the evidence against steady state models, the evidence for isotropy has been challenged, and one probably should retain some doubts about this assumption.

The evidence for spatial homogeneity is more tenuous, though belief in it is very strong, perhaps as a reaction against the centuries of the opposite view in which humanity was considered to be at the centre of the universe. The difficulty in principle is that with a finite speed of light, assertions about spatial homogeneity are statements about a region of space-time we do not observe, and so rest on assumptions about the universe's evolution. Moreover, they are assertions about the physics of indefinitely distant regions, from which we can only receive indefinitely faint signals. Apart from philosophical reasons, the main bases for assuming spatial homogeneity are the distributions of galaxies and radio sources and the lack of perturbations in the CMWBR of the kind large inhomogeneities would produce. The main difficulty with this assumption is that the largest scale on which inhomogeneities in galaxy distributions have been observed has increased and seems to be approaching the scale of the observed universe.

The FLRW models are remarkably successful in accounting for the observed features of the universe. Up to 1980, the main defect was that there had been insufficient time for spontaneous fluctuations to produce the observed departures from spatial homogeneity. That defect may be cured by the theory of inflation; in this theory, phase transitions of a scalar field (or some other fields which mimic a scalar field for this purpose) produce an effective cosmological constant leading to a period of rapid expansion during which suitable fluctuations can be formed².

Given this success, why should we consider other models further? One can advance the following reasons. First, we cannot be sure that the observations support the FLRW models unless we can show that other models cannot fit. Secondly, we would like to know if there is an explanation of the observed isotropy and homogeneity other than just assuming symmetric initial conditions; such conditions are very special in the sense that if we consider the backward evolution in time of data at some finite time after the big-bang they are unstable to perturbations. Thirdly, we should have some alternatives available in case the residual difficulties of the FLRW models prove insuperable. One may also note that models may have applications as descriptions of local phenomena or particular features of the universe (MacCallum, 1993a). In general other models are required for any situation where the FLRW models or linearized perturbations on them are an inadequate description.

The main limitations of work on the alternative geometries are that it usually assumes a perfect fluid matter content, and that only mathematically tractable models have been much explored, though we probably stand at the threshold of an era when fully four-dimensional numerical studies will be possible. There are a very large number of studies of metrics which generalize the standard (FLRW) models in the sense that they reduce to those models when certain parameters are set to zero. These families of solutions, many of which still use symmetric geometries but coupled with generaliza-

²I have elsewhere argued that other claims of inflation theory are ill-founded (MacCallum, 1987; MacCallum, 1993a). I do not want to repeat those arguments here, but I would in particular say I still do not think it has been proved that inflation accounts for the observed isotropy or that it predicts that the universe must have the critical density, despite the widespread belief in these claims.

tion of the energy-momentum content, rather than completely asymmetric geometries³, have recently been comprehensively surveyed in a mammoth review by Krasinski (as yet unpublished, but discussed in Krasinski (1990) and at the GR13 conference (1993)): Krasinski has read over 1900 papers in preparation of that review (private communication).

2. COSMOLOGICAL KINEMATICS AND DYNAMICS

The dynamics of relativistic cosmologies can be presented in a rather general setting developed mainly by Ehlers and Ellis (Ehlers, 1961; Ellis, 1971; MacCallum, 1973). One merit of this approach is that it illuminates the similarities and differences between the Newtonian and relativistic cases, though this will not be developed here.

The conventions I will use are that the metric g_{ab} has signature $(-, +, +, +)$, partial derivatives are written $u^a,_b \equiv \partial u^a / \partial x^b$, covariant derivation is denoted thus:

$$u^a,_b = u^a,_b + \Gamma^a_{cb} u^c,$$

the Ricci identity reads

$$p_{b;cd} - p_{b;cd} = R^a_{bcd} p_a,$$

the Ricci tensor and scalar are defined by

$$R_{bd} = R^a_{bad}, \quad R = g^{bd} R_{bd}$$

and the Einstein equations read

$$R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} = \kappa T_{ab}$$

or

$$R_{ab} = \kappa (T_{ab} - \frac{1}{2} T g_{ab}) + \Lambda g_{ab}.$$

The constant κ can be taken to be 1 (by setting $c = 1 = 8\pi G$ where c is the velocity of light and G is Newton's gravitational constant).

The treatment is based on a choice of a (suitably differentiable) unit timelike vector field u^a or its corresponding congruence of timelike integral curves. The velocity u^a is often considered to be that of fundamental observers in cosmology. The main problem is how to choose or define u^a in a suitable and meaningful way. Having done that, all other tensors can be decomposed into parts parallel to u^a and parts perpendicular, using the projection tensor

$$h_{ab} \equiv g_{ab} + u_a u_b.$$

In particular we can split the covariant derivative of the velocity, similarly to the methods used in Newtonian fluid dynamics, as

$$u_{a;b} = \frac{1}{3} \theta h_{ab} + \sigma_{ab} + \omega_{ab} - \dot{u}_a u_b,$$

where ω is skew and σ is symmetric and tracefree. The acceleration $\dot{u}_a \equiv u_{a;b} u^b$ is zero for geodesic motion. The vorticity ω_{ab} describes the rotation of the velocity field relative

³Although the Szekeres solutions and their generalizations, which have no Killing vectors, are included here, these models do have "intrinsic symmetries", as discussed in section 5.5 below.

to a dynamically non-rotating frame. The expansion θ describes the overall volume change, and can be used to define a length scale ℓ (up to a factor on each integral curve of u^a) such that $\theta = 3\dot{\ell}/\ell$, where the overdot denotes a directional derivative in the u^a direction. The “Hubble constant” H is $\theta/3$ (and is not in fact constant in general cosmological models). The shear σ_{ab} describes anisotropy of expansion. The interpretations of the quantities θ , ω_{ab} and σ_{ab} are obtained by considering the evolution of infinitesimal vectors connecting neighbouring curves of the timelike congruence.

One of the Einstein field equations is the Raychaudhuri equation

$$\frac{3\ddot{\ell}}{\ell} = 2\omega^2 - 2\sigma^2 + \dot{u}_{;a}^a - \frac{1}{2}(\rho + 3p) + \Lambda \quad (1)$$

where $\dot{\theta} + \theta^2/3 = 3\ddot{\ell}/\ell$ has been used and $2\omega^2 = \omega_{ab}\omega^{ab}$, $2\sigma^2 = \sigma_{ab}\sigma^{ab}$, ρ is the energy of the matter content and p is its pressure. We see that a positive Λ acts as a repulsive force, as do ω^2 and a positive $\dot{u}_{;a}^a$, the former of these being a “centrifugal” term, while the shear and positive density and pressure are attractive. This leads to Raychaudhuri’s theorem (Raychaudhuri, 1955), the first of the singularity theorems, which says that if the acceleration and vorticity vanish and expansion does not, and if Λ and $\rho + 3p$ are non-negative, then a singularity where $\ell \rightarrow 0$ occurs in a finite time, at most $\ell/\dot{\ell} = 1/H$. This provided the conceptual basis for all subsequent singularity theorems (Hawking and Ellis, 1973).

The main differences between Newtonian and relativistic behaviour can be accounted for as follows. The \dot{u} terms arise from special relativistic considerations, the “magnetic part” of the tidal force which arises in decomposition of the Riemann tensor arises just from using Riemannian geometry, and additional terms involving parts of the energy-momentum tensor are purely general-relativistic.

The formalism briefly introduced here uses a distinguished timelike vector u^a , often chosen as the velocity of matter or according to some geometric property, but does not make any choice of spatial axes. The extension which does that is called the orthonormal tetrad formalism and was developed by Ellis (1966); I have reviewed it elsewhere (MacCallum, 1973). This technique has been extensively used in cosmology, especially by members of the Waterloo group (Collins, Wainwright, Goode, Szafron, Hewitt and others). It is perhaps worth noting that I would now derive it from a treatment in terms of differential forms rather than as I did it in my 1973 article.

The other important aspect which I shall omit here is the derivation of the observable relations, which requires the integration of the geodesic equations. It is an unpleasant fact that even for comparatively simple metrics, these equations are generally impossible to integrate analytically, and their numerical integration can be a non-trivial task (see e.g. the contribution by Ribeiro in these proceedings).

3. ROBERTSON-WALKER SOLUTIONS

These are defined as the models which are everywhere isotropic, as observed by some special set of observers. Using their velocity field, the isotropy immediately implies that $\omega = \sigma = 0 = \dot{u}^a$ and that all invariantly-defined quantities, such as ρ , θ and p , have zero spatial derivatives, e.g. $h^a_b\rho_{,a} = 0$. The vanishing of ω implies the existence of spacelike surfaces of constant density orthogonal to u^a , and the Ricci identity and field equations then tell us that the spacetime is conformally flat, the energy-momentum tensor is of

perfect fluid form and there is a function t such that $u_a = -t_{,a}$ which can be taken as the time coordinate. The remaining field equations imply

$$\frac{3\ddot{\ell}}{\ell} + \frac{1}{2}(\rho + 3p) - \Lambda = 0 \quad (2)$$

$$\dot{\rho} + 3(\rho + p)\frac{\dot{\ell}}{\ell} = 0 \quad (3)$$

$$2\rho + 2\Lambda - {}^3R = 6\frac{\dot{\ell}^2}{\ell^2} \quad (4)$$

where 3R is the scalar curvature of the three-dimensional spatial surfaces of homogeneity, which, due to the isotropy, must have a Ricci tensor ${}^3R_{ab} = {}^3Rh_{ab}/3$ where $K \equiv {}^3R/6 = k/\ell^2$ for a constant k (which can be made ± 1 or 0 by choice of ℓ). (2) is Raychaudhuri's equation, (3) follows from the Bianchi identities, and (4) from the Gauss equation of the embedding of the three-dimensional spaces into the space-time. (4) is called the Friedman equation, and is the main equation to be solved. These equations are often re-written in terms of the Hubble constant H , the deceleration parameter $q \equiv -\ddot{\ell}/H^2\ell$ and the density parameter $\Omega \equiv \rho/3H^2$. Given an equation of state we can write everything in terms of Ω and $\dot{\Omega}$. The critical density $\Omega = 1$ is that for which $\Lambda = 0$ implies $k = 0$.

We can now extract some general properties of the possible types of solution. First, the only solution with $\dot{\ell} = 0$ has $(\rho + 3p)/2 = \Lambda > 0$ and $3k/\ell^2 = \rho + \Lambda > 0$. This is the Einstein static universe of 1917 and the value of Λ is the critical value Λ_c .

If $p \geq 0$ then $\rho \geq M/\ell^3$ as ℓ decreases, for some constant M . Thus at small ℓ we have $3\dot{\ell}^2 \approx \rho\ell^2$ (the increasing accuracy of this approximation as the singular point $\ell = 0$ is approached is referred to by inflation theorists as the flatness problem). For the perfect gas law $p = (\gamma - 1)\rho$ with constant γ , we have in this limit $\ell \propto t^{2/3\gamma}$, $H = 2/3\gamma$, $\rho \propto t^{-2}$. These relations are exact for all t if $\Lambda = k = 0$, in which case the solutions with $\gamma = 1$, and $\gamma = 4/3$, when the matter is described respectively as "dust" and "(incoherent) radiation", are due to Einstein and de Sitter (1917) and Tolman (1933) respectively. The solution with $\gamma = 2$ ("stiff matter") has $\ell \propto t^{1/3}$.

For large ℓ it is usually assumed that $p \rightarrow 0$ so that $\rho \approx M/\ell^3$ for a constant M . The behaviour is then dominated by Λ if it is not 0 , and otherwise by k . Solutions with $\Lambda < 0$ must re-collapse. Solutions with $\Lambda > 0$ behave like the de Sitter solution (in which $\rho = p = 0$) where $\ell \propto \exp(\sqrt{\Lambda}t/3)$ and $H = \sqrt{\Lambda}/3$. If $\Lambda = 0$ the models re-collapse if $k = 1$, are approximately Einstein-de Sitter if $k = 0$, and expand with $\ell \propto t$ if $k = -1$ (the exact solution corresponding to this last case, the Milne model, is in fact a portion of flat space in unusual coordinates).

The models all expand indefinitely (or contract indefinitely) unless it is possible for $\dot{\ell}$ to be 0 , for which we need $(\rho + \Lambda)\ell^2 = 3k$, so that for normal matter we need $k = 1$. The evolution can have a "coasting phase" where the behaviour is close to the Einstein static universe if $\Lambda \approx \Lambda_c$. To obtain more detail one can usefully plot $\rho\ell^2$ against $3k - \Lambda\ell^2$.

4. SPATIALLY HOMOGENEOUS MODELS

That a metric can be spatially homogeneous but anisotropic can be a surprise, but it is clear once one considers the possibility of anisotropic expansion. The models of this

type include as special cases the RW models⁴, and are generally known as the Bianchi models because the possible symmetries were classified by Bianchi (1897). In fact there is only one exceptional case, the Kantowski-Sachs metric (Kantowski and Sachs, 1966)

$$ds^2 = -dt^2 + A^2(t)dx^2 + B^2(t)(d\theta^2 + \sin^2 \theta d\phi^2)$$

which, it should be noted, although containing spheres where t and x are constant, is not spherically symmetric in the usual sense since there is no interior to any of the spheres: instead the topology is $S^2 \times \mathbb{R}^2$ (or some topology obtained from that by identifications).

More extensive reviews than I give here are given by Ryan and Shepley (1975), and by Rosquist and Jantzen (1988), and in my two 1979 papers (MacCallum, 1979a; MacCallum, 1979b). New papers on these solutions continue to appear, and indeed the dynamics of Bianchi models provided one of the main problems considered at the workshop.

4.1. Bianchi Classification

The Bianchi models have three independent spacelike Killing vector fields ξ_α acting on the surfaces of homogeneity, where the Greek indices range from 1 to 3. These obey

$$[\xi_\alpha, \xi_\beta] = C^\gamma{}_{\alpha\beta}\xi_\gamma$$

where the $C^\gamma{}_{\alpha\beta}$ are constants called structure constants. Bianchi classified the possible distinct sets of structure constants, and the three-dimensional positive-definite metrics with such symmetries, but did not discuss cosmology at all.

It follows from the existence of the symmetry group that there are basis vector fields e_α in each surface such that $\mathcal{L}_\xi e = 0$ for each e and ξ , and hence the scalar products between the e_α are constant in each surface. From this it follows that in each surface one can (by choice of the e_α at an initial point) ensure that

$$[e_\alpha, e_\beta] = C^\gamma{}_{\alpha\beta}e_\gamma.$$

Following the methods of Schücking and Behr as in Ellis and MacCallum (1969), we can choose the basis vectors so that their dual one-forms ω^α (this use of ω has nothing to do with the vorticity ω) obey

$$\begin{aligned} d\omega^1 &= -n_1\omega^2 \wedge \omega^3, \\ d\omega^2 &= -n_2\omega^3 \wedge \omega^1 + a\omega^1 \wedge \omega^2, \\ d\omega^3 &= -n_3\omega^1 \wedge \omega^2 + a\omega^1 \wedge \omega^3. \end{aligned}$$

The possible classes are then as shown in Table 2, where the parameter h is defined to be $a^2/n_2 n_3$. The parametrization of the structure constants can be done in other ways, and coordinates can also be introduced in various ways; the most common alternatives are the one used here (MacCallum, 1979a; Kramer *et al.*, 1980), and the one due to Taub (1951) and used by Ryan and Shepley (1975).

⁴To be specific, the $k = 1$ RW model is of Bianchi type IX, the $k = 0$ model of type I or VII₀, and the $k = -1$ model of type V or VII_h (Ellis and MacCallum, 1969; Grishchuk, 1967).

Table 2. The Bianchi classification

Class	Type	a	n_1	n_2	n_3
A	I	0	0	0	0
	II	0	0	0	1
	VI ₀	0	1	-1	0
	VII ₀	0	1	1	0
	VIII	0	1	1	-1
	IX	0	1	1	1
	V	a	0	0	0
	IV	a	0	0	1
	VI _h	a	0	1	-1
B	III = VI ₋₁	1	0	1	-1
	VII _h	a	0	1	1

4.2. Metric And Field Equations

The metrics of the Bianchi models can now be written as

$$ds^2 = -dt^2 + g_{\alpha\beta}(t)(\omega^\alpha{}_\mu dx^\mu)(\omega^\beta{}_\nu dx^\nu).$$

The time-dependence of the ω^α remains to be chosen. One alternative is to take $g_{\alpha\beta} = \text{diag}(1, 1, 1)$ (Ellis and MacCallum, 1969), the orthonormal tetrad method; here the evolution is entirely described by the ω . Another is to take time-independent ω^α , so that the time evolution is entirely in the $g_{\alpha\beta}$ (Ryan and Shepley, 1975). A perhaps even better method is to use the parametrization of the Bianchi models by automorphism group variables.

This can be briefly described as follows. Take a transformation

$$\hat{\omega}^\alpha = M^\alpha{}_\beta \omega^\beta.$$

This is an automorphism of the symmetry group if the $\{\hat{e}_\alpha\}$ obey the same commutation relations as the $\{e_\alpha\}$. The matrices M are time-dependent. This idea was present in several early treatments (Heckmann and Schücking, 1962; Collins and Hawking, 1973; Lukash, 1976) before being fully utilized (Harvey, 1979; Jantzen, 1979; Siklos, 1980; Roque and Ellis, 1985; Jaklitsch, 1987). In particular it appears in work which grew from Misner's methods for the Mixmaster case (Ryan and Shepley, 1975) but unfortunately the type IX case was highly misleading in that for Bianchi IX (and no others except Bianchi I) the rotation group is an automorphism group. The automorphism groups are given explicitly by Harvey (1979).

The advantages of this method arise because one can choose the matrices M for convenience, e.g. so that the new metric coefficients $\hat{g}_{\alpha\beta}$ become diagonal. The real dynamics is then in these metric coefficients, while the time-dependent parts of M are not true dynamical variables (see Jantzen (1979), Rosquist and Jantzen (1988)) and so the equations simplify (Siklos, 1980).

One can also choose a different time coordinate, replacing dt in the metric by $Nd\tau + N_\alpha \omega^\alpha_\mu dx^\mu$; here N is called the lapse and N_α the shift. Introducing a shift is really useful only for “tilted” cosmologies, where the matter content has a velocity not orthogonal to the surfaces of homogeneity (see e.g. King and Ellis (1973)). Lapse can however be very useful in re-casting the equations to make them more tractable or even solvable analytically; this idea has been explored systematically by Jantzen and collaborators (Jantzen, 1988; Rosquist *et al.*, 1990).

Whichever of these formalisms are used, the field equations can now be written down explicitly and investigated further. Many exact solutions are known, including some found since the review by Kramer *et al.* (1980). I shall not discuss these, but instead turn to some more general matters.

4.3. Generality

One question one can ask is how large the space of solutions is, i.e. how many parameters can be chosen arbitrarily (since the equations are ordinary differential equations, there are no free functions in the initial data)? The answer, in the form given by Siklos (1984), is set out in Table 3. In this table the numbers in brackets represent the number of parameters if h is itself considered as a parameter. The table shows why Bianchi types VIII, IX, VI and VII attract the most interest as dynamical systems (though the simplicity of Bianchi I has led to it being the favourite model for first attempts at testing effects of anisotropy in cosmology).

The results in this table have been re-analysed from the point of view of the Hamiltonian systems concerned by Ashtekar and Samuel (1991). They show that the actual number of degrees of freedom from this point of view, which is different because the equivalences used in Table 3 are not generated by the Hamiltonian constraints, is related to global topology. They also prove that Class B models cannot be closed by topological identification.

Table 3. Number of free parameters in the general solution for various Bianchi types and energy-momenta

Type	Vacuum	Perfect fluid
I	1	2
II	2	5
VI ₀	3	7
VII ₀	3	7
VIII	4	8
IX	4	8
V	1	5
IV	3	7
VI _h ($h \neq -1/9$)	3(4)	7 (8)
VII _h	3(4)	7 (8)
VI _{-1/9}	4	7

4.4. Diagonalization

Another question which was posed in various works is the characterization of the circumstances in which the metric can be brought to a purely diagonal form (i.e. using a constant automorphism matrix to make $g_{\alpha\beta}$ diagonal). The answer can be summarized as follows (MacCallum, 1972).

Class A: For all models in this class a diagonal T_{ab} in the frame given by the shear eigenvectors implies a diagonal metric, except for Bianchi types I and II. The diagonal case is still possible in those types.

Class B: The requirement that T_{ab} and $g_{\alpha\beta}$ be simultaneously diagonal implies that either

- (i) there is an anisotropic stress which violates the dominant energy condition, or
- (ii) in terms of the generic form (Ellis and MacCallum, 1969) of the decomposition of the commutators of an orthonormal tetrad, $n^\alpha_\alpha = 0$ and \mathbf{a} is an eigenvector of σ^a_b .

4.5. Lagrangians and Hamiltonians

The equations for the Bianchi models are ordinary differential equations. This raises the question of whether the usual Lagrangian for the Einstein equations can be reduced to a form $\int \mathcal{L}(g, \dot{g}, t) dt$ in terms of purely time-dependent variables g characterizing the metric (which then allows a similar form for the Hamiltonian). Initially it was believed this was so (Ryan, 1972), but by direct computation and comparison, I, and independently Estabrook and Wahlquist, noticed this was not correct. The reason, suggested by a remark of Hawking, was given by Taub and me (1972). Unfortunately the question of possible cures was incorrectly investigated in that paper; the correct version was given by Sneddon (1975). The discussion has been re-written in terms of improved sets of variables by Bogoyavlenskii (1985) and Jantzen (1984). The topic still attracts attention (though, sadly, most subsequent papers have repeated the original errors); a recent paper suggests more radical ways of finding a Hamiltonian that will work (Hojman *et al.*, 1992).

The result is that a Lagrangian always exists in Class A, and exists in Class B if $n^\alpha_\alpha = 0$ (here it is assumed that we have a Lagrangian for the matter content, so this statement concerns only the geometric part of the field equations, but see MacCallum *et al.* (1970)). Other cases require a non-holonomic constraint, or equivalently a “friction force”.

The reason for the failure of the reduction can be seen by considering the derivation of Euler-Lagrange equations from a one-dimensional Lagrangian. If we vary

$$I = \int_{x_1}^{x_2} f(x, y, y') dx$$

we obtain an expression

$$\delta I = \left[\frac{\partial f}{\partial y'} \delta y \right]_{x_1}^{x_2} + \int_{x_1}^{x_2} \delta y \left(\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \right) dx$$

so that to obtain the Euler-Lagrange equation the first term on the right must vanish. This is usually achieved by taking δy to vanish at the boundaries, but that option is not open to us in homogeneous cosmologies since δy must itself be homogeneous, and so would vanish everywhere if it vanished at spatial boundaries. The alternative is

the “natural” boundary conditions $\partial f / \partial y' = 0$. Considering these leads to the results above.

The virtue of the Lagrangian approach, when it works, is that it allows us to describe the evolution of the model as a classical potential-well problem (though with time-dependent potential). This in turn can then be approximated by simplified forms of the potentials.

4.6. Bianchi Model Equations As Dynamical Systems

What I briefly touch on here is a major topic of the workshop. In discussing the ODEs for Bianchi models as dynamical systems a number of aspects have been considered, for example

- Cases which reduce to planar systems, enabling the use of phase-plane methods, e.g. Collins (1971, 1974)
- Good choices of variables
- Compactification of the phase space
- Chaos

The phase-plane cases are well covered in the various reviews cited, and the relation to the theory of chaos is extensively discussed in other contributions, so I will make a couple of further remarks on choice of variables (see also subsection 4.2); that one has to be careful in choice and interpretation of variables is shown by Siklos’ (1991) criticisms of Barrow and Sonoda (1986). I conclude (in the next subsection) with some pointers to the literature on further choices which compactify the phase space.

The usual way to complete the decomposition of the metric, after using group automorphisms to give a diagonal $g_{\alpha\beta}$, is to parametrize this diagonal matrix with some form of the variables introduced by Misner (1969), e.g.

$$g_{\alpha\gamma} = e^{\beta_0} \text{diag}(\exp(\beta_+ + \sqrt{3}\beta_-), \exp(\beta_+ - \sqrt{3}\beta_-), \exp(-2\beta_+)).$$

Using the reduced Lagrangian for diagonal Class A (vacuum) models this leads to the potentials shown in Figure 1. Additional potentials representing the effects of anisotropic stresses, tilt and so on can be added.

4.7. Phase Space Compactification

The idea of phase space compactification for Bianchi models is already present in the phase-plane studies (see Collins (1971)), but was introduced for general cases by S.P. Novikov and Bogoyavlenskii (see Bogoyavlenskii (1985)). The process consists of the normalization of configuration variables so that physically admissible states lie within some bounded region, whose boundaries usually represent big-bang singularity states (or possibly indefinitely expanded states) and such that the dynamical system remains analytic on the physical region and its boundaries. The formulation obtained is typically then exploited by (a) finding Lyapunov functions, showing that the evolution is driven to regions near the boundaries of the phase space and (b) using analyticity, together with the behaviour of critical points and separatrices on the boundary, where the dynamical system simplifies, to derive the asymptotic behaviour. It is usually the behaviour near

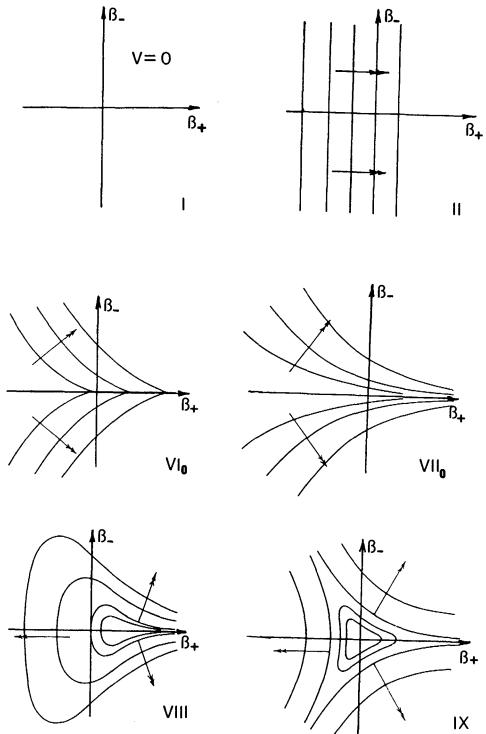


Figure 1. Equipotential contours in the $\beta_+\beta_-$ plane for evolution in the diagonal Class A vacuum Bianchi models. The double-headed arrows indicate the direction of exponential expansion.

an initial singularity that is studied, though the far-future behaviour can be treated similarly.

Three main groups have developed these treatments. Bogoyavlenskii and his colleagues, the pioneers, used a rather cumbersome set of variables and a larger system than necessary. Their studies covered all Bianchi types and included (in work of Peresetskii described in Bogoyavlenskii (1985)) tilted perfect fluid matter content. In particular this work gave a rigorous basis to the conclusions about oscillatory singularities in Bianchi IX and other Bianchi models which had been obtained by Belinskii *et al.* (1970) and Misner (1969). (This does not affect the arguments about the inhomogeneous cases, reviewed in MacCallum (1982).)

Jantzen, Rosquist and collaborators (Jantzen, 1984; Rosquist *et al.*, 1990) have coupled the automorphism variables with Hamiltonian treatments and compactification in a powerful formalism. The idea of compactifying Jantzen's variables appeared first⁵ in a paper by Rosquist (1984) and led on to the "regularization" discussed by Rosquist and Jantzen (1988).

Wainwright and colleagues (e.g. Wainwright and Hsu (1989)) have used a different, and in some respects simpler, set of automorphism variables, which are well-suited for studying asymptotic behaviour because their limiting cases are physical evolutions of simpler models rather than singular behaviours. For example, they were able to show that in diagonal Class A systems the four-dimensionally self-similar models (i.e. those with a timelike homothety in addition to the isometry group) act as attractors for the evolution far from the big bang. (All such solutions for perfect fluid were found in Hsu and Wainwright (1986) and Jantzen and Rosquist (1986) following earlier recognition of their importance by Wainwright.) The same approach has been used for the oscillatory singularities (Wainwright and Hsu, 1989), and solutions with other behaviours, e.g. some which approach plane wave solutions near the singularity, have been found also.

Similar ideas can be used for the Kantowski-Sachs models too. As well as the qualitative results, these methods have enabled new exact solutions to be found, and some general statements about the occurrence of such solutions to be made.

5. INHOMOGENEOUS MODELS

In this section I shall give a brief run-down of possible cases extracted from my paper at a recent Nato ASI school (MacCallum, 1993b).

5.1. Inhomogeneous Self-Similar Models

Any model with a homothety could be described as self-similar, and among exact solutions (e.g. those listed in Kramer *et al.* (1980)) such solutions are remarkably common (Koutras, 1992). The occurrence of Bianchi models with an additional timelike homothety as attractors of evolution has already been mentioned. Self-similar models in general have recently been reviewed by Carr (1994). The definitions of self-similarity in the literature can be confusing, because the same term is also used when there is a homothety of the spatial sections rather than the whole space-time.

⁵Following a suggestion by me.

This subsection introduces those self-similar models which are like Bianchi models but with one of the isometries replaced by a homothety. The corresponding generalization of the Kantowski-Sachs models, in the sense that they have two-dimensional surfaces of constant curvature and a homothetic motion, was discussed by Collins and Lang (1987). Since the position-dependence of the metric in the direction of the homothety is completely determinate, the field equations still reduce to ordinary differential equations. The main lines of investigation are the same as for Bianchi models. The first paper discussing these in detail was by Eardley (1974), and there has been subsequent work by Wu⁶, Luminet and others (Wu, 1981; Luminet, 1978; Koutras, 1992; Hanquin and Demaret, 1983). In this work the homothety is taken to be spacelike.

A related family where the homothety is taken to be timelike has been considered by members of the Waterloo group (Hewitt *et al.*, 1988; Hewitt *et al.*, 1991; Hewitt, 1991). The models are special sub-cases of the “G₂ solutions” with perfect fluid. In these cases the time dependence is a power-law, but the space-dependence can be monotone or periodic, the asymptotic behaviour at spatial infinity can be vacuum or a spatially homogeneous model, the periodic cases are stable against anisotropy increase, and the behaviour near the singularities can be dominated by acceleration terms in (1).

All these cases give rise to dynamical systems some of which have not yet been fully investigated. However, it should be noted that these models are often physically unsatisfactory, for example because some of them have singularities at spatial infinity.

5.2. Spherically Symmetric Models

These have a metric

$$ds^2 = -e^{2\nu} dt^2 + e^{2\lambda} dr^2 + R^2(d\theta^2 + \sin^2 \theta d\phi^2)$$

where ν , λ and R are functions of r and t . The precise functional forms in the metric depend on the choice of coordinates and the additional restrictions assumed. It should be noted that there are so few undetermined functions that a sufficiently-complicated energy-momentum will fit a totally arbitrary choice of the remaining functions: in my view this should not be regarded as a solution, since no equation is actually solved!

Some important subcases have been studied, notably:

- [1] The dust (pressureless perfect fluid) cases, originally studied by Lemaître, but usually named after Tolman and Bondi;
- [2] McVittie's 1933 solution representing a black hole in an FLRW universe;
- [3] The “Swiss cheese” model constructed by matching a Schwarzschild vacuum solution inside some sphere to an exterior FLRW universe;
- [4] Shearfree fluid solutions (Wyman, 1946; Kustaanheimo and Qvist, 1948; Stephani, 1983; McVittie, 1984);
- [5] Self-similar solutions, discussed by Carr (1994).

Spherically symmetric models, especially Tolman-Bondi, have often been used to model galactic scale inhomogeneities, in various contexts such as galaxy formation, the effects of inhomogeneity on magnitude-redshift relations, simple gravitational lenses, and the formation of primordial black holes. On a larger scale, inhomogeneous spherical

⁶Due to confusions between Chinese and European name orderings, this author is often erroneously referred to as Chao.

spacetimes have been used to model clusters of galaxies, variations in the Hubble flow due to the supercluster, the evolution of cosmic voids, the observed distribution of galaxies and simple hierarchical models of the universe. Most of this work used Tolman-Bondi models, sometimes with discontinuous density distributions; the most recent work in this area is by Ribeiro, and is reviewed by him in another paper in this proceedings.

The spherically-symmetric self-similar solutions hold a special place, due to the conjecture that just as in blast waves in Newtonian fluids, the self-similar solutions may be the endpoint of evolution. Such a model has been applied to the evolution of cosmic voids (see Bogoyavlenskii (1985) and Carr (1994)). Relevant solutions include some for black holes in FLRW, non-linear perturbations in FLRW and bubbles in the early universe.

In general spherically symmetric models give rise to systems of partial rather than ordinary differential equations, but the self-similar cases of course do lead to dynamical systems in the usual sense. However, it is my impression those systems are not chaotic.

5.3. Cylindrically Symmetric And Plane-Symmetric Models

These types of solutions, usually in the static case only, are used to model cosmic strings and domain walls. Locally the two classes can be the same solutions, differing only in their global topology, i.e. by whether or not the orbit of one of the Killing vectors has topology \mathbb{R}^4 or S^1 ; this has led to some confusion in the literature.

The usual (though not the only) form for the cylindrically symmetric metrics is

$$ds^2 = -T^2 dt^2 + R^2 dr^2 + Z^2 dz^2 + 2W dz d\phi + \Phi^2 d\phi^2$$

where T , R , Z , W and Φ depend on r (and, in the non-stationary case, t) and ϕ is periodic, and, for the plane symmetric case,

$$ds^2 = -T^2 dt^2 + R^2 dr^2 + X^2(dx^2 + x^2 d\phi^2)$$

where T , R and X are functions of r (and perhaps t). The static cases all belong in Harness's (1982) general class. The equations (in the stationary case) are ordinary and could be treated as a dynamical system, but are probably not chaotic.

5.4. G_2 Cosmologies

I use the above title as a general name for all cosmological metrics with two spacelike Killing vectors (and hence two essential variables). The cylindrical and plane metrics, and many of the Bianchi metrics, are special cases of G_2 cosmologies.

G_2 cosmologies admit a number of specializations, such as:

- [1] the Killing vectors commute;
- [2] the orbits of the G_2 are orthogonal to another set of 2-dimensional surfaces V_2 ;
- [3] the Killing vectors individually are hypersurface-orthogonal;

[4] the matter content satisfies conditions allowing generating techniques. Among the classes of metrics covered here are colliding wave models, cosmologies with superposed solitonic waves, and what I call “corrugated” cosmologies with spatial irregularities dependent on only one variable.

From the point of view of dynamical systems, these classes all give partial differential systems of a rather special type.

5.4.1. Non-abelian G_2 . The metrics where the Killing vectors do *not* commute have been very little studied: it is known they cannot admit orthogonal V_2 if the fluid flows orthogonal to the group surfaces (unless they have an extra symmetry) and that if the fluid is thus orthogonal it is non-rotating (Bugalho, 1987; Van den Bergh, 1988). So we now take only cases where the Killing vectors commute.

5.4.2. Abelian G_2 Not Orthogonally Transitive. The case without orthogonal V_2 has also been comparatively little studied, but recently some exact solutions which have one hypersurface-orthogonal Killing vector and in which the metric coefficients are separable, have been derived and studied (Van den Bergh *et al.*, 1991; Van den Bergh, 1992). Most of the solutions have singularities at finite spatial distances or can be regarded as inhomogeneous perturbations of the Bianchi VI_{-1} models.

5.4.3. Abelian G_2 Orthogonally Transitive. The cases with orthogonal V_2 were classified by Wainwright (1979;1981), and a number of specific examples with fluid energy-momentum are known (e.g. Wainwright and Goode (1980); Kramer (1984)). A recent solution found by Senovilla (1990) attracted much attention because it is non-singular (Chinea *et al.*, 1991), evading the focussing conditions in the singularity theorems by containing matter that is too diffuse: it is closely related to an earlier solution of Feinstein and Senovilla (1989), and has been generalized (Ruiz and Senovilla, 1992; Van den Bergh and Skea, 1992; S.W. Goode at GR13 (unpublished)). The metrics investigated in this class generally have Kasner-like behaviour near the singularity (though some have a plane-wave asymptotic behaviour (Wainwright, 1983)) and become self-similar or spatially homogeneous in the far future.

5.4.4. G_2 Solutions Obtainable By Generating Techniques. Finally we come to the most-studied class, those where the generating techniques are applicable. The matter content must have characteristic propagation speeds equal to the speed of light, so attention is restricted to vacuum, electromagnetic, neutrino and “stiff fluid” (or equivalently, massless scalar field with a timelike gradient) cases. However, FLRW fluid solutions can be obtained by using the same methods in higher-dimensions and using dimensional reduction. There are useful reviews covering the cosmological, cylindrical, and colliding wave sub-classes (Carmeli *et al.*, 1981; Verdaguer, 1985, 1993; Griffiths, 1991).

The metrics can be written in a form covering also the related stationary axisymmetric metrics as

$$ds^2 = \epsilon f_{AB} dx^A dx^B + \delta e^{2\gamma}((dx^4)^2 - \epsilon(dx^3)^2)/f$$

where A, B take values 1, 2 and the values of f_{AB} can be written as a matrix

$$\begin{pmatrix} f & -f\omega \\ -f\omega & f\omega^2 + \epsilon(x^3)^2/f \end{pmatrix}$$

The case $\delta = -\epsilon = 1$ gives the stationary axisymmetric metrics, the case $\delta = \epsilon = 1$ the cylindrical cases and $\epsilon = -\delta = 1$ the cosmological cases. Physically these classes differ in the timelike or spacelike nature of the surfaces of symmetry and the nature of the gradient of the determinant of the metric in those surfaces.

Some studies have focussed on the mathematics, showing how known vacuum solutions can be related by solution-generating techniques (Kitchingham, 1984), while others have concentrated on the physics of the evolution and interpretative issues (Verdaguer (1993) gives a particularly good discussion of the cases obtainable by the solitonic generating method). Over-simplifying, one can say that the models where there are no unacceptable singularities are typically Kasner-like near the singularity, and settle down to self-similar or spatially homogeneous models with superposed high-frequency gravitational waves at late times (Adams *et al.*, 1982; Carmeli and Feinstein, 1984; Feinstein, 1988). There does not appear to be any chaotic behaviour but there are interesting questions about the meaning of “soliton” in this context.

5.5. Other Models

Solutions with less symmetry than those above have been little explored. Following Krasinski one can divide the cases considered into a number of classes (in which I only mention a few important special subcases); any relation to chaos has yet to be determined.

1. The Szekeres-Szafron family (also independently found by Tomimura). These have in general no symmetries, but the spatial sections are conformally flat and individually admit isometries of three-dimensional space which do not form a four-dimensional isometry: in this sense the models have intrinsic symmetry. They contain an irrotational non-accelerating fluid. Tolman-Bondi universes are included in this class, as are the Kantowski-Sachs metrics; some generalizations are known, such as the rotating inhomogeneous model due to Stephani (1987).
2. Shearfree irrotational metrics (Barnes, 1973) which include the conformally flat fluids (Stephani 1967a, 1967b) and McVittie’s spherically symmetric metric.
3. The Vaidya-Patel-Koppar family, which represent an FLRW model containing a “Kerr” solution, using null radiation and an electromagnetic field. The physical significance of these metrics is dubious.
4. Some other special cases such as Oleson’s Petrov type N fluid solutions.

6. CONCLUDING REMARKS

The aim of this paper has been to introduce the various types of cosmological model which give the dynamical systems currently under investigation and featured in the rest of the workshop, and to indicate briefly other classes which might be referred to in discussions of the cosmological implications of the investigations of chaotic dynamics even if these models themselves are not directly the object of study.

The use of these models to illuminate issues in astrophysical cosmology has largely been ignored here, but in closing I should emphasize that a very wide range of such issues is still routinely discussed using the anisotropic and/or inhomogeneous cosmological models. Such issues include the thermodynamics of the universe, the effect of global topology, and the interpretation of the CMWBR anisotropies. Thus the present efforts to understand better the dynamics of, and possible chaos in, cosmological models may be useful in better understanding the universe as well as better understanding the mathematics of the models.

REFERENCES

- Adams, P.J., Hellings, R.W., Zimmerman, R.L., Farshoosh, H., Levine, D.I. and Zeldich, S., 1982, Inhomogeneous cosmology: gravitational radiation in Bianchi backgrounds. *Astrophys. J.*, **253**, 1.
- Ashtekar, A. and Samuel, J., (1991). Bianchi cosmologies: the role of spatial topology. *Class. Quantum Grav.* , **8**, 2191-2215.
- Barnes, A., 1973, On shearfree normal flows of a perfect fluid. *Gen. Rel. Grav.*, **4**, 105.
- Barrow, J.D. and Sonoda, D.H., 1986, Asymptotic stability of Bianchi type universes. *Phys. Reports*, **139**, 1-49.
- Belinskii, V.A., Lifshitz, E.M. and Khalatnikov, I.M., 1970, The oscillatory regime of approach to an initial singularity in relativistic cosmology. *Usp. Fiz. Nauk.*, **102**, 463. Translated in *Sov. Phys. Uspekhi*, **13**, 475 (1971). Also see *Adv. Phys.* **19**, 525 (1970).
- Bianchi, L., 1897, Sugli spazii a tre dimensioni che ammettono un gruppo continuo di movimenti. *Mem. di Mat. Soc. Ital. Sci.*, **11**, 267. Reprinted in Opere, vol. IX, ed. A. Maxia (Rome: Edizioni Cremonese) 1952.
- Bogoyavlenskii, O.I., 1985, *Methods of the qualitative theory of dynamical systems in astrophysics and gas dynamics*. (Berlin and Heidelberg: Springer-Verlag) [Russian original published by Nauka, Moscow, 1980].
- Bugalho, M.H., 1987, Orthogonality transitivity and cosmologies with a non-Abelian two-parameter isometry group. *Class. Quantum Grav.*, **4**, 1043.
- Carmeli, M., Charach, C. and Malin, S., 1981, Survey of cosmological models with gravitational, scalar and electromagnetic waves. *Phys. Repts.*, **76**, 79.
- Carmeli, M. and Feinstein, A., 1984, The cosmological peeling-off property of gravity. *Phys. Lett. A*, **103**, 318-320.
- Carr, B.J., 1994, Self-similar perturbations of the Friedman universe and large-scale structure. To appear in *Class. Quantum Grav.*

- Chinea, F.J., Fernandez-Jambrina, F. and Senovilla, J.M.M., 1991, A singularity-free space-time. *Phys. Rev. D*, **45**, 481-6.
- Coley, A.A. and Tupper, B.O.J., 1983, Zero-curvature Friedman-Robertson-Walker models as exact viscous magnetohydrodynamic cosmologies. *Astrophys. J.*, **271**, 1.
- Collins, C.B., 1971, More qualitative cosmology. *Commun. math. phys.*, **23**, 137.
- , 1974, Tilting at cosmological singularities. *Commun. Math. Phys.*, **39**, 131.
- Collins, C.B. and Hawking, S.W., 1973, Why is the universe isotropic? *Astrophys. J.*, **180**, 37.
- Collins, C.B. and Lang, J.M., 1987, A class of self-similar space-times and a generalization. *Class. Quantum Grav.*, **4**, 61-78.
- Eardley, D.M., 1974, Self-similar spacetimes: geometry and dynamics. *Commun. math. phys.*, **37**, 287.
- Ehlers, J., 1961, Beiträge zur relativistischen Mechanik kontinuierlicher Medien (Contributions to the relativistic mechanics of continuous media). *Akad. Wiss. Lit. Mainz, Abh. Math.-Nat. Kl.* (11). (English translation by P.K.S. Dunsby, University of Capetown, 1993).
- Ellis, G.F.R., 1966, *Geometry, relativity and cosmology*. Adams Prize Essay, University of Cambridge.
- , 1971, Relativistic cosmology. In Sachs, R.K., editor, *General relativity and cosmology*, volume XLVII of *Proceedings of the International School of Physics "Enrico Fermi"* (New York and London: Academic Press), pp. 104–182.
- Ellis, G.F.R. and MacCallum, M.A.H., 1969, A class of homogeneous cosmological models. *Commun. math. phys.*, **12**, 108–141.
- Ellis, G.F.R. and Stoeger, W., 1987, The ‘fitting problem’ in cosmology. *Class. Quantum Grav.*, **4**, 1697–1729.
- Feinstein, A., 1988, Late-time behaviour of primordial gravitational waves in expanding universe. *Gen. Rel. Grav.*, **20**, 183–190.
- Feinstein, A. and Senovilla, J.M.M., 1989, A new inhomogeneous cosmological perfect fluid solution with $p = \rho/3$. *Class. Quantum Grav.*, **6**, L89–91.
- Ferrari, V., 1990, Colliding waves in general relativity. *General Relativity and Gravitation, 1989*, ed. Ashby, N., Bartlett, D.F. and Wyss, W. (Cambridge, New York and Melbourne: Cambridge University Press), pp. 3–20.
- Griffiths, J.B., 1991, *Colliding plane waves in general relativity*. Oxford mathematical monographs. (Oxford: Oxford University Press).
- Grishchuk, L.P., 1967, Cosmological models and spatial-homogeneity criteria. *Astr. Zh.*, **44**, 1097–1103. (In Russian: English translation in *Soviet Astronomy — A.J.*, **11**, 881–5 (1968)).
- Hanquin, J.-L. and Demaret, J., 1983, Gowdy $S^1 \times S^2$ and S^3 inhomogeneous cosmological models. *J. Phys. A*, **16**, L5.
- Harness, R.S., 1982, Spacetimes homogeneous on a timelike hypersurface. *J. Phys. A*, **15**, 135.
- Harvey, A.L., 1979, Automorphisms of the Bianchi model Lie groups. *J. Math. Phys.*, **20**, 251.
- Hawking, S.W. and Ellis, G.F.R., 1973, *The large-scale structure of space-time*. (Cambridge: Cambridge University Press).

- Heckmann, O. and Schücking, E., 1962, Relativistic cosmology. *Gravitation: an introduction to current research*, ed. Witten, L. (New York and London: J. Wiley).
- Hewitt, C.G., 1991, An investigation of the dynamical evolution of a class of Bianchi VI_{1/9} cosmological models. *Gen. Rel. Grav.*, **23**, 691–712.
- Hewitt, C.G., Wainwright, J. and Glaum, M., 1991, Qualitative analysis of a class of inhomogeneous self-similar cosmological models II. *Class. Quantum Grav.*, **8**, 1505–1518.
- Hewitt, C.G., Wainwright, J. and Goode, S.W., 1988, Qualitative analysis of a class of inhomogeneous self-similar cosmological models. *Class. Quantum Grav.*, **5**, 1313–1328.
- Hojman, S.A., Nunez, D. and M.P. Ryan, jr., 1992, *A minisuperspace example of non-Lagrangian quantization*. *Phys. Rev. D*, **45**, 3523–7.
- Hsu, L. and Wainwright, J., 1986, Self-similar spatially homogeneous cosmologies: orthogonal perfect fluid and vacuum solutions. *Class. Quantum Grav.*, **3**, 1105–1124.
- Jaklitsch, M.J., 1987, *First order field equations for Bianchi types II-VI_h*. University of Capetown preprint 87-6.
- Jantzen, R.T., 1979, The dynamical degrees of freedom in spatially homogeneous cosmology. *Commun. math. phys.*, **64**, 211.
- , 1984, Spatially homogeneous dynamics: a unified picture. *Cosmology of the early universe*, ed. Ruffini, R. and Fang, L.-Z. (Singapore: World Scientific), pp. 233–305. Also in *Gamow cosmology*, (Proceedings of the International School of Physics 'Enrico Fermi', Course LXXXVI), 1987, ed. Ruffini, R. and Melchiorri, F. (Amsterdam: North Holland), pp. 61–147.
- , 1988, Power law time lapse gauges. *Phys. Rev. D*, **37**, 3472.
- Jantzen, R.T. and Rosquist, K., 1986, Exact power law metrics in cosmology. *Class. Quantum Grav.*, **3**, 281.
- Kantowski, R. and Sachs, R.K., 1966, Some spatially homogeneous cosmological models. *J. Math. Phys.*, **7**, 443.
- King, A.R. and Ellis, G.F.R., 1973, Tilted homogeneous cosmological models. *Commun. math. phys.*, **31**, 209.
- Kitchingham, D.W., 1984, The use of generating techniques for space-times with two non-null commuting Killing vectors in vacuum and stiff perfect fluid cosmological models. *Class. Quantum Grav.*, **1**, 677–694.
- Koutras, A., 1992, *Mathematical properties of homothetic space-times*. Ph.D. thesis, Queen Mary and Westfield College.
- Kramer, D., 1984, A new inhomogeneous cosmological model in general relativity. *Class. Quantum Grav.*, **1**, 611–618.
- Kramer, D., Stephani, H., MacCallum, M.A.H. and Herlt, E., 1980, *Exact solutions of Einstein's field equations*. (Berlin: Deutscher Verlag der Wissenschaften, and Cambridge: Cambridge University Press). (Russian translation: "Tochnie resheniya uravnenii Einsteina", 418 pp., translated by I.V. Mitskovich, V.D. Zakharov and S.V. Rumyantsev and edited by Yu. S. Vladimirov (Moscow: Energoisdat) 1982).
- Krasinski, A., 1990, Early inhomogeneous cosmological models in Einstein's theory. *Modern cosmology in retrospect*, ed. Bertotti, B., Bergia, S., Balbinot, R. and Messina, A. (Cambridge: Cambridge University Press).
- Kustaanheimo, P. and Qvist, B., 1948, A note on some general solutions of the Einstein field equations in a spherically symmetric world. *Comment. Math. Helsingf.*, **13**, 12.

- Lukash, V.N., 1976, Physical interpretation of homogeneous cosmological models. *Nuovo Cim. B*, **35**, 268.
- Luminet, J., 1978, Spatially homothetic cosmological models. *Gen. Rel. Grav.*, **9**, 673.
- MacCallum, M.A.H., 1972, On ‘diagonal’ Bianchi cosmologies. *Phys. Lett. A*, **40**, 385–6.
- , 1973, Cosmological models from the geometric point of view. *Cargese Lectures in Physics*, vol. 6, ed. Schatzman, E., (New York: Gordon and Breach), pp. 61–174.
- , 1979a, Anisotropic and inhomogeneous relativistic cosmologies. *General relativity: an Einstein centenary survey*, ed. Hawking, S.W. and Israel, W., (Cambridge: Cambridge University Press), pp. 533–580. (Russian translation: *Obshchaya teoria otnositel’nosti* ed. Ya. A. Smorodinskii and V.B. Braginskii, (Moscow: Mir), 1983. Also reprinted in *The early universe: reprints*, ed. E.W. Kolb and M.S. Turner, (Reading, Mass: Addison-Wesley), pp. 179–236, 1988).
- , 1979b, The mathematics of anisotropic spatially-homogeneous cosmologies. *Physics of the expanding universe*, ed. Demianski, M. volume 109 of *Lecture Notes in Physics*, (Berlin and Heidelberg: Springer-Verlag) pp. 1–59.
- , 1982, Relativistic cosmology for astrophysicists. *Origin and evolution of the galaxies*, ed. de Sabbata, V. (Singapore: World Scientific), pp. 9–33. Also, in revised form, in *Origin and evolution of the galaxies*, ed. Jones, B.J.T. and J.E., Nato Advanced Study Institute Series, **97**, pp. 9–39, (Dordrecht: D.Reidel and Co.) 1983.
- , 1984, Exact solutions in cosmology. *Solutions of Einstein’s equations: techniques and results (Retzbach, Germany, 1983)*, ed. Hoenselaers, C. and Dietz, W. *Lecture Notes in Physics*, **205**, pp. 334–366, (Berlin and Heidelberg: Springer-Verlag).
- , 1987, Strengths and weaknesses of cosmological big-bang theory. *Theory and observational limits in cosmology*, ed. W.R. Stoeger, S.J., (Vatican City: Specola Vaticana), pp. 121–142.
- , 1993a, Anisotropic and inhomogeneous cosmologies. *The Renaissance of General Relativity and Cosmology (A survey to celebrate the 65th birthday of Dennis Sciama)*, ed. Ellis, G., Lanza, A. and Miller, J.C. (Cambridge: Cambridge University Press), pp. 213–233.
- , 1993b, Inhomogeneous and anisotropic cosmologies. *The origin of structure in the universe*, ed. Gunzig, E. and Nardone, P., NATO ASI Series C, **393**, (Dordrecht: Kluwer), pp. 131–159.
- MacCallum, M.A.H., Stewart, J.M. and Schmidt, B.G., 1970, Anisotropic stresses in homogeneous cosmologies. *Commun. math. phys.*, **17**, 343–7.
- MacCallum, M.A.H. and Taub, A.H., 1972, Variational principles and spatially homogeneous universes, including rotation. *Commun. math. phys.*, **25**, 173–189.
- McVittie, G.C., 1984, Elliptic functions in spherically symmetric solutions of Einstein’s equations. *Ann. Inst. Henri Poincaré*, **40**, 235–271.
- Misner, C.W., 1969, The Mixmaster universe. *Phys. Rev. Lett.*, **22**, 1071.
- Raychaudhuri, A.K., 1955, Relativistic Cosmology I. *Phys. Rev.*, **98**, 1123.
- Roque, W.L. and Ellis, G.F.R., 1985, The automorphism group and field equations for Bianchi universes. *Galaxies, axisymmetric systems and relativity: essays presented to W.B. Bonnor on his 65th birthday*, ed. MacCallum, M.A.H., (Cambridge: Cambridge University Press), pp. 54–73.
- Rosquist, K., 1984, Regularized field equations for Bianchi type VI spatially homogeneous cosmology. *Class. Quantum Grav.*, **1**, 81–94.

- Rosquist, K. and Jantzen, R.T., 1988, Unified regularization of Bianchi cosmology. *Phys. Repts.*, **166**, 89.
- Rosquist, K., Uggla, C. and Jantzen, R.T., 1990, Extended dynamics and symmetries in perfect fluid Bianchi cosmologies. *Class. Quantum Grav.*, **7**, 625–637.
- Ruiz, E. and Senovilla, J.M.M., 1992, A general class of inhomogeneous perfect-fluid solutions. *Phys. Rev. D*, **45**, 1995–2005.
- Ryan, M.P., 1972, Hamiltonian cosmology, *Lecture Notes in Physics*, **13**, (Berlin and Heidelberg: Springer-Verlag).
- Ryan, M.P. and Shepley, L.C., 1975, *Homogeneous relativistic cosmologies*. (Princeton: Princeton University Press).
- Senovilla, J.M.M., 1990, New class of inhomogeneous cosmological perfect-fluid solutions without big-bang singularity. *Phys. Rev. Lett.*, **64**, 2219–2221.
- Siklos, S.T.C., 1980, Field equations for spatially homogeneous spacetimes. *Phys. Lett. A*, **76**, 19.
- , 1984, Einstein's equations and some cosmological solutions. *Relativistic astrophysics and cosmology (Proceedings of the XIVth GIFT International Seminar on Theoretical Physics)*, ed. Fustero, X. and Verdaguer, E., (Singapore: World Scientific), pp. 201–248.
- , 1991, Stability of spatially homogeneous plane wave spacetimes I. *Class. Quantum Grav.*, **8**, 1587–1602.
- Smoot, G.F., et al ., 1992, Structure in the COBE DMR first year maps. *Astrophys. J.*, **396**, 1.
- Sneddon, G.E., 1975, Hamiltonian cosmology: a further investigation. *J. Phys. A*, **9**, 229.
- Stephani, H., 1967a, Konform flache Gravitationsfelder. *Commun. math. phys.*, **5**, 337.
- , 1967b, Über Lösungen der Einsteinschen Feldgleichungen, die sich in einen fünf-dimensionalen flachen Raum einbetten lassen. *Commun. math. phys.*, **4**, 137.
- , 1983, A new interior solution of Einstein's field equations for a spherically symmetric perfect fluid in shearfree motion. *J. Phys. A*, **16**, 3529–3532.
- , 1987, Some perfect fluid solutions of Einstein's field equations without symmetries. *Class. Quantum Grav.*, **4**, 125.
- Taub, A.H., 1951, Empty space-times admitting a three-parameter group of motions. *Ann. Math.*, **53**, 472.
- Van den Bergh, N., 1988, A class of inhomogeneous cosmological models with separable metrics. *Class. Quantum Grav.*, **5**, 167–178.
- , 1992, A qualitative discussion of the Wils inhomogeneous stiff fluid cosmologies. *Class. Quantum Grav.*, **9**, 2297–2307.
- Van den Bergh, N. and Skea, J.E.F., 1992, Inhomogeneous perfect fluid cosmologies. *Class. Quantum Grav.*, **9**, 527–532.
- Van den Bergh, N., Wils, P. and Castagnino, M., 1991, Inhomogeneous cosmological models of Wainwright class A1. *Class. Quantum Grav.*, **8**, 947–954.
- Verdaguer, E., 1985, Solitons and the generation of new cosmological solutions. *Observational and theoretical aspects of relativistic astrophysics and cosmology (Proceedings of the 1984 Santander School)*, ed. Sanz, J.L. and Goicoechea, L.J. (Singapore: World Scientific), pp. 311–350.
- , 1993, Soliton solutions in spacetimes with two spacelike Killing vectors. *Phys. Repts.*, **229**, 1–80.

- Wainwright, J., 1979, A classification scheme for non-rotating inhomogeneous cosmologies. *J. Phys. A*, **12**, 2015.
- , 1981, Exact spatially inhomogeneous cosmologies. *J. Phys. A*, **14**, 1131.
- , 1983, A spatially homogeneous cosmological model with plane-wave singularity. *Phys. Lett.*, **A99**, 301.
- Wainwright, J. and Goode, S.W., 1980, Some exact inhomogeneous cosmologies with equation of state $p = \gamma\mu$. *Phys. Rev. D*, **22**, 1906.
- Wainwright, J. and Hsu, L., 1989, A dynamical systems approach to Bianchi cosmologies: orthogonal models of Class A. *Class. Quantum Grav.*, **6**, 1409–1431.
- Wu, Z.-C., 1981, Self-similar cosmological models. *Gen. Rel. Grav.*, **13**, 625. Also see J. China Univ. Sci. Tech. **11**(2), 25 and **11**(3), 20.
- Wyman, M., 1946, Equations of state for radially symmetric distributions of matter. *Phys. Rev.*, **70**, 396.

HOMOCLINIC CHAOS IN RELATIVISTIC COSMOLOGY

Esteban Calzetta¹

IAFE, cc 67, suc 28 (1428), Buenos Aires, Argentina
and Physics Department, University of Buenos Aires

Abstract. We give a short review of Homoclinic Chaos, drawing instances of it from General Relativity, and putting emphasis on the application of Melnikov's method for the detection of this kind of behavior. We describe in some detail two concrete manifestations of Homoclinic Chaos in relativistic Cosmology, one where the Universe may be described as a Hamiltonian dynamical system, and other where this is not possible, due to the presence of viscous matter. The overall implications of Chaos for Cosmology and General Relativity are discussed.

1. INTRODUCTION

Our objective in these talks is to describe a particular manifestation of chaotic behavior, the so-called Homoclinic Chaos [1, 2, 3], to present some techniques for its detection, and to demonstrate its incidence in relativistic problems generally, and in Relativistic Cosmology in particular. For our purposes, "chaos" denotes deterministic behavior which displays random features in some well defined sense; while most chaotic phenomena are rather complex if seen over long time scales, complication alone is not sufficient qualification for chaos. Conversely, even extremely chaotic behavior may appear rather tame on short time scales, a point to which we shall return in time.

The setting of these lectures excuses me from extolling the relevance of a proper "chaos awareness" for the future development of our discipline. As subtler means of observation become available, both in Cosmology and Astrophysics, we dare explore the behavior of matter under more extreme conditions, be it in the neighborhood of collapsed objects or in the wake of a cosmic singularity. In either of these settings,

¹e-mail: calzetta@iafe.edu.ar

strong gravitational fields predominate, and the full non linearity of Einstein's equations comes into play. In this regime, the search for exact solutions or for reliable perturbative techniques may well be illusory. However, the recognition of chaotic behavior as such, and the deployment of analytical tools borrowed from Dynamical Systems Theory, can yield the necessary understanding of the structure of these phenomena, and lead us to the meaningful questions to ask of the observational data.

Both dynamical systems common sense, and experience with the mathematically simpler Yang - Mills theory [4], lead us to believe that chaos is indeed generic in strong field gravitational physics. However, to date only a handful of problems have been analyzed in any detail. While the in depth analysis of such problems as Bianchi IX cosmologies, geodesic flux in spaces of negative sectional curvatures, and motion of test particles around one or two black holes, all of them well represented in this same volume, have yielded valuable insights on the peculiarities of relativistic chaos, the mere number of these applications stands in stark contrast with the above cited expectation of universality in chaotic behavior. To bridge this gap, and either confirm or disprove this belief, a new standpoint and new techniques are needed.

Again, the setting of this workshop in oil-rich Alberta provides us with an apt analogy. In getting oil out of the ground, a complex technology of well-drilling is necessary. But drilling is expensive, and time consuming. Therefore, in searching for oil on vast and mostly unknown territories, one does not go drilling every so many feet, but employs instead some cheaper, easier to use prospector tools, like seismic information and so on. These prospector techniques never yield certainty, and so, once a promising area has been targeted, it is nevertheless necessary to go and drill to ascertain whether oil is there or not. But the enormous gain in screening barren areas, plus the added advantages of knowing what to expect when on the ground, clearly outweighs the costs of prospecting.

Using this image as a metaphor for our problem, the brilliant (but mostly custom made for their immediate object) analytical discussions of Bianchi IX cosmology in the literature, as well as the sophisticated numerical machinery now available, play the role of drilling technology. In these lectures, I shall discuss a different approach, namely the Melnikov method [1], which, I will contend, fills the specifications of a prospector tool. Let us say from the outset that, as could be expected from a prospector tool, the Melnikov method will prove easy to use, will have a wide range of applications, and will yield no certainty of actual chaotic behavior. However, a combination of wide search techniques, like Melnikov's, and narrow beam in depth analysis (mostly by numerical means), offers our best hope of establishing the actual incidence of Chaos in General Relativity. I shall present below a simple example of how such a two-step approach may be carried to success.

Actually, Melnikov's method is designed to detect only a specific form of chaotic behavior, namely Homoclinic Chaos, which occurs when an integrable Hamiltonian system is subject to a time - periodic perturbation. While we shall give below a detailed description of Homoclinic Chaos, let us say here that when it occurs, the smooth flow proper to the unperturbed dynamics becomes chaotic under the perturbation, in certain regions of phase space known as stochastic layers. When the perturbation is strong enough, different layers merge, and a stochastic sea is formed [5, 6, 7]. Within the stochastic region, the flow becomes essentially random, being equivalent to a Bernoulli Shift [8, 9]. Melnikov's method provides a test of the formation of a stochastic layer, valid to first order in the strength of the perturbation. For finite perturbations, the

results from Melnikov's method require validation from some independent, non perturbative approach.

In actual applications, the scenario most frequently leading to Homoclinic Chaos consists of a dynamical system whose dynamics can be split into an integrable, Hamiltonian part, and a (non necessarily Hamiltonian) perturbation. Time - periodicity follows from the dependence of the perturbation on the angle variables of the integrable part. To obtain Homoclinic Chaos, it is necessary that the Fourier decomposition of the perturbation in terms of angle variables contains at least two components, one of which is in resonance with the quasi periodic integrable motion. This resonance destroys one of the Kolmogorov - Arnold - Moser tori of the integrable system, and creates a separatrix instead[10]. Chaos is produced by the action of the secondary perturbations on the separatrix. The Melnikov method consists in computing a certain integral along the separatrix, which measures the success of the secondary perturbations in creating a stochastic layer[6, 7].

As we shall argue below, from the presentation of simple examples, this scenario is widespread in relativistic systems, and particularly in relativistic cosmology. The Melnikov method is apt to search for Chaos in these areas, not least because it can be applied to systems of any number of dimensions[2, 11], and because it can handle non Hamiltonian as well as Hamiltonian perturbations[12]. Of course, a problem involving the gravitational field should in principle be regarded as having an infinite number of degrees of freedom; however, oftentimes it is allowed to impose symmetries on the model, which reduce the degrees of freedom to a finite number. The configuration spaces for these reduced systems are known as Mini Super Spaces[13], and must obey certain consistency conditions[14]; the simplest Mini Super Spaces are those describing Friedmann - Robertson - Walker cosmologies, where all the geometric information about the Universe is reduced to a single variable. The dynamical behavior of multi dimensional systems being qualitatively different from low dimensional ones, it is nevertheless necessary to be able to handle Mini Super Spaces of arbitrary complexity.

As for dissipative (non - Hamiltonian) perturbations, they arise because of energy and angular momentum loss due to gravitational radiation, or because of conventional dissipative processes, like viscosity, in models with fluid matter. In cosmological models, the first source of dissipation may be disposed of by assuming a closed Universe, as we shall do in these lectures, following Quantum Cosmological prejudices. However, the second one stands, and it is likely to remain in any realistic cosmological model. In this context, it is interesting to observe that even quantized fields behave as a dissipative fluid in the context of the Very Early Universe[15].

In summary, our contention in these talks is that Homoclinic Chaos is indeed a common occurrence in relativistic dynamics, and that a combination of perturbative methods such as Melnikov's, to target suitable candidates, and high power numerical techniques, to in depth studies of those singled out models, is possibly our best strategy to investigate its many fold instances. These techniques therefore set the stage for the next and most important step, which is to deploy our chaos savvy to address the traditional concerns of General Relativity and Cosmology.

Let us describe briefly the contents of these talks. In next section, we shall give a swift introduction to Homoclinic Chaos. We shall present a measure preserving linear automorphisms on a torus ("Arnold's Cat Map") as the basic model bearing the intuitive idea of what a chaotic system looks like. In particular, we shall give a non technical discussion of Kolmogorov Entropy [16], and show how a positive Kolmogorov Entropy is

associated to randomness in the long time behavior of the system. We shall then proceed to consider near integrable Hamiltonian systems, to exhibit invariant subsets where the restricted dynamics is close to Arnold's Cat Map. After giving a concrete example of one such system (a test particle orbiting a Schwarzschild Black Hole, perturbed by gravitational waves [18]), we shall discuss briefly the generality of these models.

In the third, and main, section of this paper, we shall show how Homoclinic Chaos is present already in an extremely simple cosmological model, namely, a spatially closed Friedmann - Robertson - Walker (FRW) Universe filled with a conformally coupled but massive scalar field [19]. As discussed above, we shall employ Melnikov's method to detect signatures of chaos when the mass is small, and then we shall proceed to present hard evidence of chaos through non perturbative (numerical) methods. This simple example will be used as background for a general discussion of the likelihood and relevance of Chaos in our Universe.

While in this example the Universe may be described as a Hamiltonian system, even after the perturbation is switched on, the Melnikov method may be applied with equal ease to the case of non Hamiltonian perturbations [1, 12]. In the fourth section of this paper, we shall take advantage of this fact to discuss a more realistic example of chaotic FRW cosmology, whereby besides the scalar field, a cosmological fluid is also present. Moreover, we shall allow this fluid to display bulk viscosity (the only viscous term allowed by FRW symmetry). Assuming a very simple constitutive relation, we shall show that Melnikov's method suggests an enhancement of the chaotic behavior of the model, due to viscosity. However, these results should be seen as preliminary, in so far as the necessary corroborating numerical work is still in progress[20].

We shall therefore present three concrete instances of Homoclinic Chaos in General Relativity. The conclusions to be drawn from these examples are briefly considered in our final remarks.

2. HOMOCLINIC CHAOS

In this section, we shall discuss the generalities of Homoclinic Chaos. For this, it is convenient to present from the outset a class of relatively simple dynamical systems, to provide us with a concrete image of what a chaotic system looks like. We shall also discuss briefly the concept of Kolmogorov Entropy, as central to the understanding of the random features of long time chaotic evolution.

While interesting in their own right, these models will be too simple to be directly relevant to relativistic Chaos. However, we shall show that similar behavior is found in certain invariant sets in the phase space of near integrable Hamiltonian systems. We shall demonstrate a simple perturbative procedure to detect such sets, the so-called Melnikov method. We shall then present our first relativistic example, a test particle orbiting a Schwarzschild Black Hole, whose orbit becomes chaotic under gravitational perturbations of the metric. The section closes with a brief discussion of where Homoclinic Chaos is likely to be found.

2.1. Simple Chaotic Models

As announced in the Introduction, we shall choose, as mental picture of a chaotic system, a point particle moving in discrete time on the surface of a torus $T = S^1 \times S^1$. The position of the particle at time $n + 1$ will be a linear function of its position at

times n , $\vec{\theta}_{n+1} = A\vec{\theta}_n$, where $\vec{\theta} = (\theta_1, \theta_2)$ are two angles defining a coordinate system on the torus in the usual way. We shall assume the evolution preserves areas ($\det A = 1$), and that the eigenvalues of A have the form $\lambda^{\pm 1}$, where λ is real and larger than 1. Thus the evolution shall be stretching in one direction \vec{v}_+ , and contracting in another \vec{v}_- . These two directions are not necessarily orthogonal, as A is not symmetric[16, 21].

The effect of each iteration is most easily seen if $\lambda \gg 1$. It is like opening the torus into a square, stretching it into a ribbon, and then wrapping it tightly around the torus again, keeping one point fixed. If we open up the torus one more time, it will look like we have transformed a set of thin stripes, along the \vec{v}_- direction, covering the torus, into a set of stripes along the \vec{v}_+ direction. A highly idealized version of this situation works with just two stripes: we have $\theta'_1 = 2\theta_1$, $\theta'_2 = \theta_2/2$ if $\theta_1 \leq \pi$, $\theta'_1 = 2\theta_1 - 2\pi$, $\theta'_2 = \theta_2/2 + \pi$, otherwise. This map is known as the Baker's Transformation; it is supposed to resemble a baker working on a piece of dough[22].

We claim that our discrete time dynamics is highly chaotic, meaning that the long time evolution of our particle is essentially random. We stress that this does not mean merely that the position of the particle after a long time will be very hard to compute. It means that if we partition the torus into a finite number of tiles, number them, and record the string of numbers picked up by the particle through a large series of iterations, the result will be impossible to tell, by any statistical test, from a string of like symbols generated by a suitable roulette [8]. Technically, this follows from the fact that these systems have positive Kolmogorov Entropy S_K ; in fact, we find $S_K = \ln \lambda$ [16, 23]. Let us then discuss briefly the physical meaning of S_K .

Our tiling of the torus may be regarded as a coarse measurement of the particle's position. If the j th tile T_j has area μ_j , and we ignore entirely the position of the particle, then to predict the result of this measurement we require $s_0 = -\sum_j \mu_j \ln \mu_j$ bits of information [24]. If we knew that the particle was in tile T_k after the last iteration, the information needed to predict the next measurement is instead $s_1^k = -\sum_j \mu_{jk} \ln \mu_{jk}$, where μ_{jk} is the ratio of the area of $T_j \cap A(T_k)$ to μ_k . Thus, to predict the outcome of the second measurement, given the result of the first, we need in the average $s_1 = \sum_k \mu_k s_1^k$ bits of information. In general, $s_1 \leq s_0$; the more we know, the easier to predict the next result. More generally, if we call s_n the information needed to predict the outcome of the $n + 1$ -th measurement, knowing the n previous results, then s_n will be a non increasing sequence, and will have a finite limit s_∞ . If we consider the values of s_∞ for all possible finite tilings, and select the greatest, we obtain the Kolmogorov Entropy[16, 23].

Thus, the statement that a system has a strictly positive S_K , means that no matter how many times we perform a certain coarse - grained measurement, we shall never be able to predict the next outcome; an irreducible stochastic element is present in the long time dynamics. Of course, once a particular tiling leading to a positive s_∞ is found, further refinement of the initial tiling will only lead to a higher asymptotic entropy. Conversely, for an integrable system, all tilings lead to $s_\infty = 0$. In the example of the Baker's transformation, the split of the square into two tiles, $\theta_1 \leq \pi$ and $\theta_1 \geq \pi$, already realizes the maximum entropy, $S_K = \ln 2$ [16].

We see thus how randomness and Chaos are present in even rather simple systems. Maybe more surprising, we shall see now how certain invariant sets in perturbed Hamiltonian systems actually behave in the same way that our discrete time evolution on the torus. This will pave the way for the presentation of our first example of Chaos in relativistic dynamics.

2.2. Homoclinic Chaos

In this section, we shall discuss Homoclinic Chaos in Hamiltonian systems, still keeping the proceedings at a somewhat abstract level. To avoid unnecessary complications, though, let us restrict ourselves from the beginning to the consideration of a one dimensional point particle moving inside a potential well $V(x)$ (as we remarked in the Introduction, the restriction to one dimensional systems is not essential). Such a system is necessarily integrable, the Hamiltonian $H_0 = p^2/2m + V(x)$ providing a constant of motion.

Let us assume that the potential has an unstable equilibrium point, say, at $x = 0$, that is, we have $V(0) = E_s$, $V'(0) = 0$, $V''(0) < 0$, where a prime stands for a derivative with respect to x . Moreover, we shall assume that there is a separatrix, that is, a trajectory with energy E_s , approaching asymptotically to 0 as $t \rightarrow \pm\infty$. For $E \leq E_s$, the motion will be periodic, with the orbits approaching the separatrix uniformly as $E \rightarrow E_s$, and the frequency ω of the periodic motion going to zero in that limit. For concreteness, we shall assume that for $E \sim E_s$, we may approximate [7]

$$\omega(E) \sim \frac{\omega_0}{\ln \frac{aE_s}{(E_s - E)}} \quad (1)$$

Which usually holds in these problems. Here ω_0 is some reference frequency, and a is a numerical constant.

We shall subject now our particle to a time periodic external perturbation, and study under what conditions the separatrix is thereby substituted by an stochastic layer. To do this, let us focus first on a particle executing periodic motion below the separatrix. In the absence of perturbation, the value E of H_0 would be a constant of motion, and the particle would return to its initial position after a period $2\pi/\omega(E)$. In this lapse, then, the phase $\phi = \nu t$ of the perturbation will have increased by $2\pi\nu/\omega(E)$. Under the perturbation, H_0 is not precisely constant, and the change in ϕ is not quite as before; however, for a weak enough perturbation, we may still write

$$E' = E + \Delta E; \quad \phi' = \phi + 2\pi \frac{\nu}{\omega(E')} \quad (2)$$

for the energy and phase after the particle has gone once through a complete orbit.

The energy shift ΔE may be found by integrating dH_0/dt over the trajectory. Since the trajectory, however, remains uniformly close to the separatrix, for small perturbations we can compute this integral on the separatrix itself. In this approximation, the change in E will depend only on the initial phase ϕ , and the map Eq. (2) will preserve areas, irrespective of whether the perturbation is Hamiltonian or not. It is then called the "separatrix map"[5, 6, 7].

To obtain a better insight on the nature of the separatrix map, let us look for fixed points. A fixed point (E_f, ϕ_f) should obey $\omega(E_f) = \nu/k$, which always has solutions for k large enough, since $\omega \rightarrow 0$ as we approach the separatrix, and $\Delta E = 0$. Thus, if ΔE , which in general shall be a periodic function of ϕ , remains bounded away from zero, there will be no fixed points, and therefore no chance of reproducing chaotic behavior of the kind we saw in the previous section, anywhere near the separatrix.

Let us assume that ΔE vanishes at some value ϕ_0 of the phase, and let (E_k, ϕ_0) be a sequence of fixed points, with $k \rightarrow \infty$. The separatrix map close to each fixed point can be approximated by the corresponding linearized map, whose eigenvalues are

$$\lambda_k^\pm = (1 + \lambda_k) \pm \sqrt{(1 + \lambda_k)^2 - 1} \quad (3)$$

where

$$\lambda_k = -k \left(\frac{d\Delta E}{d\phi}(\phi_0) \right) \left(\frac{d \ln \omega}{dE} \right)(E_k). \quad (4)$$

Observe that, for large k , $\lambda_k \sim \nu [d\Delta E/d\phi]/\omega_0(E_s - E_k)$. Thus, unless $d\Delta E/d\phi$ vanishes at ϕ_0 , all fixed points are strongly hyperbolic close to the separatrix. We deduce then a second requirement for chaos, namely, that ΔE must actually cross zero for at least one value of the initial phase.

Finally, it can be seen that the stretching direction is essentially parallel to the ϕ axis, while the contracting direction forms a finite angle with it. We thus reproduce, in a neighborhood of each fixed point, essentially the same behavior of the linear automorphisms of the torus we studied in the previous section. Namely, the separatrix map, suitable restricted to these neighborhoods, will have positive Kolmogorov Entropy $S_K \sim \ln \lambda_k$. More generally, we find highly irregular behavior in a full measure set surrounding the separatrix. We can estimate the width δE of this “stochastic layer” as the distance to the separatrix where the eigenvalues of the linearized separatrix map become of order unity. This yields

$$\delta E \sim \frac{\nu}{\omega_0} \frac{d\Delta E}{d\phi}(\phi_0). \quad (5)$$

To conclude, we find that the key to the formation of a stochastic layer is that the energy shift along the separatrix ΔE must oscillate through zero for some values of the initial phase. ΔE is given by the integral[1]

$$\Delta E = \int_{-\infty}^{+\infty} dt \left\{ \frac{\partial H_0}{\partial p} \delta \dot{p} + \frac{\partial H_0}{\partial x} \delta \dot{x} \right\} \quad (6)$$

where $\delta \dot{p}$ and $\delta \dot{x}$ stand for the variation in the equations of motion, and the integral is performed along the unperturbed separatrix.

The Melnikov method consists precisely in computing this “Melnikov integral” and verifying the conditions above. It is a crucial element of the method that, to compute the Melnikov integral, only the unperturbed separatrix must be known, and some analytical form of the perturbation. It is therefore unnecessary to solve the perturbed dynamics, or even, for that matter, the unperturbed motion away from the separatrix. This feature is the key to the simplicity and flexibility of the Melnikov method in actual applications.

If the perturbation itself is Hamiltonian, $\delta \dot{x} = \partial \delta H / \partial p$, $\delta \dot{p} = -\partial \delta H / \partial x$, and the Melnikov integral reduces to

$$\Delta E = \int_{-\infty}^{+\infty} dt \left\{ \delta H, H_0 \right\} \quad (7)$$

where the brackets are the usual Poisson ones.

2.3. An Example from General Relativity

At this point in our discussion, it is convenient to pause and consider our first example of relativistic chaos. Our system shall be a test particle of mass μ orbiting a Schwarzschild black hole of mass M , and being perturbed by gravitational radiation. The unperturbed motion shall be the geodesic flow in the Schwarzschild background

$$ds^2 = -(1 - \frac{2M}{r})dt^2 + (1 - \frac{2M}{r})^{-1}dr^2 + r^2(d\theta^2 + \sin^2 \theta d\phi^2); \quad (8)$$

without loss of generality, we may assume that the orbit lies in the equatorial plane $\theta = \pi/2$.

The role of Hamiltonian is played by $-p_t$, the momentum canonically conjugated to time. Since angular momentum L is also conserved, it is possible to find an effective Hamiltonian describing the radial motion. This Hamiltonian has fixed points, describing circular orbits, provided $12M^2\mu^2/L^2 < 1$. The circular orbit closer to the horizon is unstable; if moreover $12M^2\mu^2/L^2 \geq 3/4$, there shall be a separatrix. The separatrix describes an orbit slowly peeling off the unstable circular one, then making one large excursion away from the hole, bouncing back from the Newtonian tail of the potential, and winding back around the unstable circular orbit. Let us bring in the perturbations. For this, we shall follow Regge and Wheeler's classification of gravitational waves on a Schwarzschild background, restricting ourselves, for simplicity, to waves leaving invariant the equatorial plane [17]. By solving the constraint $p^2 = -\mu^2$ in the perturbed metric, it is possible to read off the new terms in $-p_t$, which constitute the perturbation δH , and write down the Melnikov integral. While in this case we do not have a closed analytic expression for the perturbation, the high frequency approximations given by Regge and Wheeler and others are enough to show that, in the high frequency limit, the Melnikov integral does indeed change sign at certain values of the initial phase. We thereby are allowed to conclude that, for weak waves, the separatrix is destroyed, and a stochastic layer appears in its place [18].

While this result has some relevance to the study of the motion of real particles near Black Holes, what we would like to stress here is the economy of means by which the Melnikov method allows us to establish this highly nontrivial result.

After this relativistic rest, let us return for a short while to our previous abstract discussion.

2.4. Where to Look for Homoclinic Chaos

The example above may seem contrived in that it assumes a peculiar structure in the unperturbed phase space (the unstable fixed point and its separatrix), and the perturbation was brought in, so to speak, "from the outside". A much more common scenario leading to Homoclinic Chaos depends upon internal resonances in systems with two or more degrees of freedom. We shall discuss here the two dimensional case; the generalization to higher dimensions is immediate[6, 7].

Let us consider a system described by periodic canonical coordinates $\vec{\theta} = (\theta_1, \theta_2)$ and their canonically conjugated momenta $\vec{I} = (I_1, I_2)$. Let the Hamiltonian be $H = H_0(\vec{I}) - V(\vec{I}, \vec{\theta})$, where the perturbation V has a Fourier decomposition

$$V(\vec{I}, \vec{\theta}) = \sum_{\vec{n}} V_{\vec{n}}(\vec{I}) \cos\{\vec{n} \cdot \vec{\theta} - \phi_{\vec{n}}\} \quad (9)$$

where $\vec{n} = (n_1, n_2)$. Observe that the "unperturbed" Hamiltonian H_0 is already in action - angle form.

As it is well known, this system is still integrable, to first order in the perturbation, unless on the resonant lines

$$\vec{n} \cdot \vec{\omega} = 0, \quad (10)$$

where $\vec{\omega} = (\omega_1, \omega_2) = \nabla_{\vec{I}} H_0$ are the frequencies of angular motion. Therefore, each term in V will be dynamically important mostly when \vec{I} is close to the corresponding resonance. Moreover, the resonant motion will still conserve H_0 to first order, since

close to a resonance we find $(\vec{I}) \cdot \sim \vec{n}$. Thus, for a given value $H_0 = E$ preordained, we may approximate Eq. (9) by

$$V(\vec{I}, \vec{\theta}) = \sum_{\vec{n}} V_{\vec{n}} \cos\{\vec{n} \cdot \vec{\theta} - \phi_{\vec{n}}\}, \quad (11)$$

where $V_{\vec{n}} = V_{\vec{n}}(\vec{I}_{\vec{n}})$, and $I_{\vec{n}}$ is the common solution to $H_0(\vec{I}) = E$ and Eq. (10).

Let us choose a vector \vec{k} , and study the motion in a neighborhood of $\vec{I}_{\vec{k}}$. Let us assume, for simplicity, that $k_2 \neq 0$, $\omega_1 = \omega_1(\vec{I}_{\vec{k}}) \neq 0$, and $\mu^{-1} = k^i k^j \partial_i \partial_j H_0(\vec{I}_{\vec{k}}) \neq 0$. We may introduce new action variables L and J through $I^1 = I_{\vec{k}}^1 + Jk^1 + L$, $I^2 = I_{\vec{k}}^2 + Jk^2$. Their conjugated angles are $\tau = \theta_1$ and $\Theta = \vec{k} \cdot \vec{\theta} - \phi_{\vec{k}}$, respectively. Singling out the resonant terms in the Hamiltonian, we get

$$H \sim E + \omega_1 L + (1/2\mu)J^2 + c_1 LJ + c_2 L^2 - V_{\vec{k}} \cos \Theta - \delta V, \quad (12)$$

where $c_{1,2}$ are constants, and

$$\delta V = \sum_{\vec{n} \neq \vec{k}} V_{\vec{n}} \cos\{\vec{n} \cdot \vec{\theta} - \phi_{\vec{n}}\} = \sum_{\vec{n} \neq \vec{k}} V_{\vec{n}} \cos\{(1/k_2)[n_2 \Theta + (n_1 k_2 - n_2 k_1)\tau - \varphi_{\vec{n}}]\}, \quad (13)$$

where $\varphi_{\vec{n}} = k_2 \phi_{\vec{n}} - n_2 \phi_{\vec{k}}$.

We see from Eq. (12) that, for weak perturbations, τ still evolves linearly in time, $\tau \sim \omega_1 t$. It is convenient to take τ , rather than t , as parameter. The evolution of Θ as τ unfolds is described by the Hamiltonian

$$H_r = -L \sim \left(\frac{1}{2\omega_1 \mu}\right) J^2 + \left(\frac{V_{\vec{k}}}{\omega_1}\right) [1 - \cos \Theta] - \left(\frac{\delta V}{\omega_1}\right) \quad (14)$$

where we have added an unessential constant.

The reduced Hamiltonian obviously describes a nonlinear pendulum subject to time periodic perturbations. The “unperturbed” Hamiltonian H_{r0} (i. e., the pendulum) has a stable fixed point $\Theta = 0$, corresponding to $H_{r0} = 0$, and an unstable one $\Theta = \pm\pi$, corresponding to $H_{r0} = E_s = 2V_{\vec{k}}/\omega_1$. There is a separatrix joining the unstable point to itself; if this separatrix is destroyed by the perturbations, then Homoclinic Chaos will follow.

In keeping with the approximations above, to study the fate of the separatrix we consider only the strongest perturbation. This one is provided by the closest resonance, corresponding to $\vec{k}_1 = (k_1 - 1, k_2)$. The calculation of the Melnikov integral is standard, and yields (cfr. Section (2.2))

$$\Delta E = 16\pi\omega_1\mu \left(\frac{V_{\vec{k}_1}}{V_{\vec{k}}}\right) \frac{e^{\Gamma/2}}{\sinh \Gamma} \sin \phi, \quad (15)$$

where $\Gamma = \pi\omega_1\sqrt{(\mu/V_{\vec{k}})}$. We see that ΔE displays isolated zeroes, and therefore we may conclude, from Melnikov’s method, that a stochastic layer is formed, its width given by Eq. (5). Observe that, while our argument has been perturbative, ΔE , and therefore also the width of the stochastic layer, vanish to all orders in $V_{\vec{k}}$. This highlights the need for non perturbative, independent proofs of the presence of chaotic behavior in actual applications.

The stochastic layer extends along the unperturbed homoclinic loop, which in turn wanders off the original resonant line by an amount $J_{max} \sim 2\sqrt{\mu V_{\vec{k}}}$. When this

shift in the resonance is large enough, the stochastic layers from neighboring resonances may overlap. As shown by Chirikov [5], resonance overlapping marks the formation of a stochastic sea out of the merging individual layers.

The example developed in this section indicates that Homoclinic Chaos is a rather generic feature of near integrable Hamiltonian systems of several degrees of freedom. With it in mind, we have the necessary theoretical baggage to search for classical chaotic cosmological models. We turn now our attention to this, the real focus of these lectures.

3. HOMOCLINIC CHAOS IN HAMILTONIAN COSMOLOGY

In this section, we shall profit of the discussion above to present what may be the simplest conceivable chaotic cosmology, consisting of a spatially closed FRW Universe coupled to a massive scalar field. This model, which may be described as a Hamiltonian system, is integrable in the massless limit. We shall use Melnikov's method to show the formation of stochastic layers to first order in the mass of the field. This result is then corroborated by a numerical calculation of the Poincaré Sections of the dynamics. The possible relevance of cosmological chaos is discussed.

3.1. Simple Hamiltonian Cosmological Models

As it is well known, General Relativity can be formulated as a constrained Hamiltonian system. The canonical coordinates are the induced geometry and matter fields on an arbitrary Cauchy surface, while their momenta are related to the extrinsic curvature and the normal derivatives of the matter fields. The constraints are infinite in number, and are associated with the invariances of the theory, which allow for arbitrary deformations of the original surface [25]. In this case, the Hamiltonian is actually identically zero, but among the infinite number of constraints we find the "Hamiltonian" ones, which generate evolution in "many fingered time" [26]. The configuration space for General Relativity is known as Super Space [27].

This picture is geometrically satisfactory, but usually too involved to be used in building actual cosmological models. Moreover, it is a fact that, when seen on the largest accessible scales, our Universe displays striking symmetry properties, which would not be expected in a generic solution to Einstein's equations. Thus, the natural departing point for a study of the dynamical properties of cosmological models is not Super Space, but rather restricted configuration spaces, or Mini Super Spaces (MSS) [13]. Given a restriction procedure, we may obtain a MSS action by restricting the Einstein - Hilbert action functional to geometries and matter configurations within the MSS, and then find the MSS equations of motion by a variational principle, or else we may apply the restriction procedure to Einstein's equations themselves. For an acceptable MSS, these procedures should be equivalent, which usually hinges on whether it is possible to keep suitable boundary conditions, while taking variations within the MSS. Restriction procedures based on imposing that geometry and fields be invariant under a simple Lie group of transformations (such as rotations, etc.) lead to acceptable MSS's [14].

An important symmetry assumption to be imposed on possible cosmological models is that of homogeneity, that is, that there should be a three dimensional Lie group of isometries, so that the orbits of the group define a preferred spacelike slicing of the Universe, and the group acts freely on each orbit. Under this assumption, the different "Hamiltonian Constraints" merge into a single Hamiltonian, generating the evolution in

the normal direction to the preferred Cauchy slices. Moreover, if the spacelike sections are compact, then this Hamiltonian must vanish, reflecting the cancellation between gravitational potential energy, and all other forms of energy. We thus obtain the classical Wheeler - DeWitt equation.

The different homogeneous MSS's may be classified into nine inequivalent classes (the Bianchi types), according to the value of the structure constants of the isometry group [28, 29]. Among those models allowing for compact spatial sections, the richest and most interesting are those belonging to the IXth class, whose structure constants are those of SU(2). A Bianchi type IX model is described at any time by three parameters, two of them defining the shape of the corresponding Cauchy surface, and a global scale factor. Subclasses of the Bianchi IX models are the Taub models, which are axisymmetric, and the spatially closed Friedmann - Robertson - Walker (FRW) models, which are isotropic, beyond being homogeneous. In FRW cosmology, the geometry of the Universe is totally fixed, and only the global scale factor a and the matter fields play a dynamical role. Bianchi IX models have played a most important role in the development of the subject matter of this meeting; in these lectures, we shall restrict ourselves to closed FRW cosmology. The reader will find references to the original literature elsewhere in this volume.

We should point out that there is impressive evidence that our Universe is largely isotropic. Indeed, recent observations of the cosmic microwave background set limits for cosmic anisotropies at the time of decoupling much below what should be expected out of statistical fluctuations in the Early Universe alone, spurring the development of Inflationary cosmologies[30]. Evidence of homogeneity is much harder to come by, but a model which is isotropic around every point must also be homogeneous, and therefore this hypothesis seems to follow from the Copernican principle, according to which we should not assume to be living at a special location in space. A closed Universe is at best marginally consistent with observational evidence, the issue depending on whether there is enough matter to generate the necessary gravitational attraction to "close" the Universe. However, it should be remembered that estimates of the total matter density in the Universe have grown in a few years from a few percents to almost eighty percent of the required amount, according to certain observations, so the inquest can not be considered closed[31]. Moreover, a closed Universe is certainly favoured by our current understanding of Quantum Cosmology, a point we shall touch again further on[32].

Let us now describe a simple MSS model, which would retain some of the main observational features of our own Universe, as discussed above, and still lead to interesting dynamical behavior. In the first place, we shall build our model out of the FRW class, assuming from the beginning both isotropy and homogeneity. We shall not consider here more "exotic" alternatives, such as topologically nontrivial homogeneous spacetimes and so forth.

Since an autonomous, one degree of freedom Hamiltonian system is necessarily integrable, chaotic FRW models must include matter fields. The simplest case is that of real scalar fields, allowing only one polarization state; in FRW cosmology, moreover, scalar fields may be used to mimic fields of higher spin, including gravitons [33], so there is no great loss of generality in restricting ourselves to them. The simplest scalar fields are the "free" ones, that is, those described by linear equations of motion.

Cosmological scalar fields with spontaneous symmetry breaking, that is, having their ground state at non vanishing field configurations, have been extensively studied in connection with the Inflation hypothesis[34, 35]. In fact, even a free field may be used

to power inflation, provided “slow roll over” conditions hold, i. e., that the evolution of the field is slow in terms of the Hubble time[36]. Here we shall not assume this kind of condition, but rather study the free field dynamics in full generality.

A free field theory in a FRW spacetime is described by two parameters, the mass m of the field, and the coupling constant ξ , which couples the field amplitude Φ directly to the Ricci scalar R . The values $m = 0$, $\xi = (1/6)$ (in four dimensional space time) are exceptional; when both hold, the change of variables $\Phi = \phi/a$ decouples the field from the geometry. In this case, the motion is necessarily integrable.

We shall consider models where $\xi = 1/6$, but allow for a non vanishing mass. We shall also adopt a coordinate system where the metric takes the form

$$ds^2 = a^2(\eta)[-d\eta^2 + d\chi^2 + \sin^2 \chi(d\theta^2 + \sin^2 \theta d\varphi^2)]. \quad (16)$$

We shall refer to η as “conformal time”, and otherwise follow MTW conventions [25]. The energy density of the field is

$$\rho = (1/2a^4)[\phi'^2 + (1 + m^2a^2)\phi^2], \quad (17)$$

where $\phi' = d\phi/d\eta$, the field being necessarily homogeneous. The pressure is

$$p = (1/6a^4)[\phi'^2 + (1 - m^2a^2)\phi^2]. \quad (18)$$

In the massless case, $p = (1/3)\rho$, as could be expected of radiation. Due to FRW symmetry, these two quantities determine the energy momentum tensor, and therefore Einstein’s equations. The dynamics is described by the Raychauduri equation

$$a'' + a - m^2\phi^2a = 0, \quad (19)$$

and the Klein - Gordon equation

$$\phi'' + [1 + m^2a^2]\phi = 0. \quad (20)$$

Eqs. (19) and (20) derive from the Lagrangian

$$L = (1/2)[-(a'^2 - a^2) + (\phi'^2 - \phi^2) - m^2a^2\phi^2]. \quad (21)$$

Introducing the conjugated momenta $p_a = -a'$ and $p_\phi = \phi'$, we obtain the Hamiltonian

$$H = (1/2)[(p_\phi^2 + \phi^2) - (p_a^2 + a^2) + m^2a^2\phi^2]. \quad (22)$$

The Hamiltonian constraint, which is also the 00 Einstein’s equation, reads

$$H = 0. \quad (23)$$

These equations describe the simplest conceivable cosmological model which is not obviously integrable, and still retains some observational relevance. The question is whether this model is, in fact, chaotic. We proceed now to employ our “prospector tools”, introduced in the previous section, in search for an answer.

3.2. Chaotic FRW Cosmology - Perturbation Theory

According to previous discussion, we shall perform in this section a purely analytical study of our model; after discussing the main features of a typical solution, we shall split it into an integrable part and a perturbation. Then we shall use Melnikov's method to test for the presence of stochastic layers in the perturbed phase flow. In next section, we shall analize the same model by non perturbative, numerical means, in order to confirm our findings[19].

Let us begin by pointing out two elementary features of the solutions of our model, the first one being that motion is bounded. This may not seem obvious since, as shown by Eq. (18), the pressure may become negative, thus leading to "inflationary" episodes. However, these inflationary runs are stabilized by a dynamical mechanism; as the radius of the Universe increases, so does the frequency of oscillations of the Klein - Gordon field, and the pressure is brought again to positive values, thus stopping inflation. Now, boundedness of motion implies that a generic solution starts and ends at a cosmic singularity, $a = 0$. A remarkable feature of our model is that neither the equations of motion Eqs. (19) and (20) nor the Hamiltonian Eq. (22) actually become singular there. Therefore, at least formally, the solutions may be extended beyond the cosmic singularities. In this sense, a solution to our model is actually a sequence of "cosmic cycles", each describing the birth and death of an Universe. We shall adopt this view, discussing at some later point whether our conclusions will be relevant to an observer confined to a single cycle.

The second observation to be made is that, for the greatest part of a cosmic cycle, the field oscillations are faster than the rhythm of expansion and contraction of the Universe. Thus we may see the field as a "fast" variable adjusting adiabatically to the drift in the "slow" variable a . Accordingly, we would introduce an adiabatic amplitude and phase j and φ

$$\phi = \sqrt{\frac{2j}{\omega}} \sin \varphi, \quad p_\phi = \sqrt{2\omega j} \cos \varphi, \quad (24)$$

where $\omega^2 = 1 + m^2 a^2$. To preserve the Hamiltonian structure of the model, we introduce also a new geometrodynamical momentum P

$$p_a = P + \frac{m^2 a j}{2(1 + m^2 a^2)} \sin 2\varphi. \quad (25)$$

In the new variables (a, P, φ, j) , the Hamiltonian reads $H = -(H_0 + \delta H)$, where

$$H_0 = \left(\frac{1}{2}\right)[P^2 + a^2] - j\sqrt{1 + m^2 a^2} \quad (26)$$

$$\delta H = \frac{m^2 a P j}{2(1 + m^2 a^2)} \sin 2\varphi + \left[\frac{m^2 a j}{4(1 + m^2 a^2)}\right]^2 (1 - \cos 4\varphi). \quad (27)$$

There are several reasons why it is convenient to take H_0 as our unperturbed Hamiltonian, rather than simply setting $m^2 = 0$ in Eq. (22) to this end. First and foremost, the "perturbation" δH is actually small both for large and small Universes, which gives better prospects of using successfully perturbation theory. On a more formal level, H_0 clearly generates compact motions, making it more transparent to use arguments from Dynamical Systems theory. Finally, the massless Hamiltonian has degenerate frequencies, and is therefore pathological from the standpoint of Kolmogorov

- Arnold - Moser theory. As we shall see presently, this degeneracy is largely lifted in H_0 .

The “unperturbed” Hamiltonian H_0 is obviously integrable, since H_0 itself and j are constants of motion in involution. Moreover, $a = 0$ is a fixed point; it is stable if $m^2j \leq 1$, and unstable otherwise. In this second case, there is a separatrix associated with it. However, this orbit does not satisfy the Hamiltonian constraint; rather, we have $H_0 = -j$ on the separatrix.

The equations of motion are simpler in terms of a new variable $X = \omega = \sqrt{1 + m^2a^2}$, and its conjugated momentum

$$P_X = \frac{XP}{m\sqrt{X^2 - 1}}. \quad (28)$$

The unperturbed Hamiltonian becomes

$$H_0 = \left[\frac{m^2(X^2 - 1)}{2X^2} \right] P_X^2 + \left(\frac{1}{2m^2} \right) [(X - m^2j)^2 - (m^2j)^2 - 1]. \quad (29)$$

From its definition, it is clear that we must have $X \geq 1$; if $H_0 = h \geq -j$, this lower bound is actually attained. Observe that P_X diverges there; however, the time derivatives of X remain finite. In particular, we find $X' \sim 2m\sqrt{h + j}\sqrt{X - 1}$ when $X \rightarrow 1$. On the other end, the oscillations in X are bounded by an ordinary turning point X_T . Observe that in terms of the a variable, a full oscillation corresponds to the life and death of two Universes, one of them with $a \leq 0$; thus a complete oscillation actually corresponds to X going twice from 1 to X_T and back.

For this choice of variables, the motion can be explicitly integrated. However, this involves the use of elliptic integrals, and the solution is hardly transparent. For our purposes, it is more convenient to concentrate on a particular range in phase space, whereby certain simplifications will be available. A suitable range is given by the condition $m^2j \geq 1$, holding $H_0 \sim 0$ and m fixed.

Let us call K and α , respectively, the action and angle variables associated to the X oscillations. The condition that P_X vanishes at the upper turning point X_T implies that $H_0 = (1/2m^2)[(X_T - m^2j)^2 - (m^2j)^2 - 1]$. We must then have the identity

$$\frac{\partial H_0}{\partial K} = \frac{2\pi}{T(K, j)} = \frac{(X_T - m^2j)}{m^2} \frac{\partial X_T}{\partial K} \quad (30)$$

where T is the period of the X oscillation. Now, when m^2j is large, so is X_T , and $X \geq 1$ over most of the orbit, leading to $T = 4\int dX/X' \sim 4\pi$. Using this in Eq. (30), we immediately find $X_T \sim m^2j + m\sqrt{K + f(j)}$.

The independence of the period with respect to the amplitude of oscillations suggests that the motion is nearly harmonic. We therefore parametrize, for $X \gg 1$, $X = m^2j + m\sqrt{K + f(j)} \sin 2\alpha$, $P_X = m^{-1}\sqrt{K + f(j)} \cos 2\alpha$. Since we must have $K = (2/\pi)\int P_X dX$, then, $f \sim 0$. The Hamiltonian now reads

$$H_0 = \left(\frac{1}{2} \right) [K - (mj)^2 - \left(\frac{1}{m^2} \right)]. \quad (31)$$

Since the expression of X and P_X in terms of K and α depends explicitly on j , to define a new set of canonical coordinates we must also shift the field angle φ . The new angle, canonically conjugated to j , is

$$\theta = \varphi - m\sqrt{K} \cos 2\alpha. \quad (32)$$

At long last, we have expressed our problem in a form suitable for the application of the methods of Subsection (2.4).

For simplicity, let us consider only the first term in Eq. (27). Also observe that, $m\sqrt{K}$ being large in the range of interest, φ , given as a function of θ and α by Eq. (32), is a strongly oscillating function of the latter. Thus, the relatively slowly oscillating prefactor may be substituted by its mean value, which is easily seen to be of order $j \ln X_T/\pi$. The Fourier expansion of the oscillatory factor itself is straightforward [37]

$$\sin 2(\theta + m\sqrt{K} \cos 2\alpha) = J_0(2m\sqrt{K}) \cos(2\theta - \pi/2) + \sum_{n=1}^{\infty} J_n(2m\sqrt{K}) [\cos \delta_n^+ + \cos \delta_n^-] \quad (33)$$

where J_n is the usual Bessel function, and $\delta_n^{\pm} = (2\theta \pm 2n(\alpha + (\pi/4)) - (\pi/2))$. Recall that our unperturbed Hamiltonian is actually $-H_0$, and also the perturbation is $-\delta H$.

Let us identify $\vec{I} = (K, j)$. We are interested only in resonances $\vec{n} = (2n, 2)$, the resonance condition being $n - 2m^2j = 0$. For the physical value $H_0 = 0$ of the Hamiltonian, we find a sequence of resonances $\vec{I}_{\vec{n}} = ((n/2m)^2, n/2m^2)$, in the notation of subsection (2.4). The perturbation coefficients, which involve Bessel functions of large argument and order, may be evaluated at resonance, to yield

$$V_n = \left(\frac{j_n}{\pi}\right)(\ln X_T) J_n(2m\sqrt{K_n}) \sim \left(\frac{\epsilon}{m^2}\right) n^{2/3} (\ln n) \quad (34)$$

where $\epsilon \sim 0.111827\dots$ is a numerical constant.

Comparison of our model with the generic resonant Hamiltonian of Section (2.4) is easier if $n = 4k + 1$ for some k . Then the resonant term in the Hamiltonian, and its closest neighbor, become $-V_n \cos(2(\theta + n\alpha)) + V_{n-1} \sin(2(\theta + (n-1)\alpha))$. We introduce new action variables (notice the sign change with respect to Section (2.4)) $K = K_n + 2nJ - L$, $j = j_n + 2J$, and new angles $\tau = -\alpha$, $\Theta = 2(\theta + n\alpha)$. If we adopt the angle τ as “time”, then the dynamics of the (J, Θ) canonical pair shall be generated by the reduced Hamiltonian

$$-L = 4m^2 J^2 - 2V_n \cos \Theta + 2V_{n-1} \sin(\Theta + 2\tau) \quad (35)$$

which clearly reduces to the generic Hamiltonian Eq. (14), after the identification $\omega_1 = (1/2)$, $\mu^{-1} = 4m^2$. The Melnikov integral, therefore, shall be given by Eq. (15), where Γ is given by $\pi/\sqrt{4\epsilon n^{2/3} \ln n}$, and the prefactor becomes $2\pi/m^2$. We conclude that indeed stochastic layers form. Moreover, they have a constant width, of order $1/m^2$, and a constant spacing of the same order. Since, at the same time, the maximum displacement from the resonance along the separatrix grows as $\sqrt{\mu V_k} \sim n^{1/3}/m^2$, these stochastic layers overlap, and a stochastic sea arises.

We have reached the end of our perturbative foray, and our hunt has been successful, as far as perturbative methods go. The relevant question now is whether perturbation theory deserves our trust. To seek an answer, we shall proceed to cross examine our findings, re-analyzing the same model by non-perturbative, numerical means.

3.3. Chaotic FRW Cosmology - Numerics

As we have stressed in the Introduction, the best strategy in searching for instances of cosmological chaos is to combine both analytical and numerical approaches, as each will often compensate for the worst weaknesses of the other. We have just gone through a typical example of perturbative analysis of a presumably chaotic model; our positive conclusion, though, could be challenged on the grounds that it only holds for asymptotically vanishing masses. To obtain reliable results for finite masses, we must adopt a non perturbative approach, which means, in practice, a numerical one.

We shall proceed now to develop such an analysis. To make our numerical tests of the model totally independent from the perturbative arguments above, we shall revert to the “physical” variables (a, ϕ, p_a, p_ϕ) , whose evolution is generated by the Hamiltonian Eq. (22). We shall solve this evolution, integrating the equations of motion with a Runge - Kutta 5th order routine [38] , implemented on an IBM compatible 486 PC. It is well known that a method of this kind always violates the Hamiltonian constraint by some amount; an alternative is to enforce this constraint, by using it to reduce the number of dynamical variables. We have chosen instead to solve the full system of Hamilton’s equations, only monitoring the value of the Hamiltonian as one more check on the accuracy of the solutions. Runs were interrupted when the Hamiltonian exceeded a preset threshold.

A typical solution of the system conforms to the expectation of finding fast oscillations of the field modulated by the slow expansion and contraction of the Universe. “Inflationary runs” are also a common feature, but, as expected, the Universe is never able to actually avoid recollapse. While these effects are clearly visible, it is hard to the naked eye to perceive whether the motion is actually chaotic, or it is simply an harmonic motion with a large number of frequencies. To tell the difference between chaotic and complicated motions, we need a more refined approach.

Probably the deepest probe into chaos, or, at least, non integrability, concerns the topology of the orbits. For an integrable system, orbits will be confined to smooth many folds, topologically equivalent to tori [10]. This criterium of integrability is obviously independent of the parametrization of the orbits, and therefore also gauge invariant. While it is not necessary that physical criteria of chaos be gauge invariant, as a physical criterium oftentimes refers or assumes a specific observation made on the system, which defines a preferred reference frame, gauge invariant criteria nevertheless are always of interest, as those which go deepest and most directly to the essence of the problem. Means to probe directly the topology of the orbits in phase space are therefore highly desirable.

A simple way to visualize orbits in higher dimensional spaces is to take Poincaré sections[21]. In our case, we shall pick the $\phi = 0$ plane, and mark the intersection points between this plane and a given orbit, the orbit crossing the plane in a given sense. In this way, we obtain a one dimensional trace of the orbit on the plane, which will fit into a smooth curve if and only if the system is integrable.

Fig. (1)[19] shows the Poincaré sections for orbits beginning from $a = 0$ and $\phi = 0$, and $0.225 \leq p_\phi \leq 1.975$; the initial value of p_a is equal to $-p_\phi$, as follows from the Hamiltonian constraint, and the requirement that $a' > 0$ initially. We choose $m = 0.8$, and record the values of a and p_a along the orbit when $\phi = 0$ and $p_\phi \geq 0$. Low values of the momenta correspond to low values of $m^2 K$ and $m^2 j$, in the notation of the previous subsection; we expect, from our perturbative analysis, that a stochastic

sea forms when these parameters become of order one, corresponding to an initial value of $|p_a| \geq 5/\sqrt{8} \sim 5/3$. Indeed, orbits with $|p_a| \leq 1$ remain bounded all along (cfr. the blob in the lower left corner of Fig. (1), while orbits beginning with $|p_a| \geq 1$ burst and seem to fill the whole of phase space. This result brilliantly confirms our expectations; if anything, our estimate of the border of chaos has been too conservative.

Fig. (2) [19] presents a blow up of the lower left corner of Fig. (1). The conditions of the “experiment” are the same, but now the initial values of p_ϕ run only up to 0.975, and only the sector $0 \leq a \leq 1.8$, $0 \leq p_a \leq 1.5$ is shown. We can now resolve the structure of the blob, and thereby gain a valuable insight on the origin of chaos in this model. Indeed, several near resonances are clearly seen; however, these resonances do not merge, and motion remains regular. Resonance overlap occurs for the first time near $|p_a| \sim 1$; at this point, the stochastic sea is formed, and regularity is destroyed.

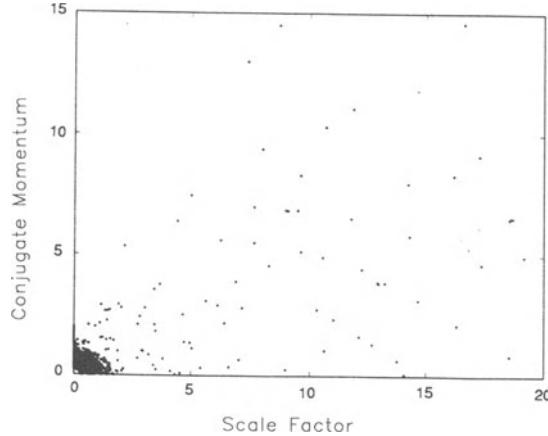


Figure 1. Poincaré sections on the $\phi = 0$ plane; in this case $m = 0.8$, and initially $a = \phi = 0$, $0.225 \leq p_\phi \leq 1.975$. Only the first quadrant is shown: $0 \leq a \leq 20$, $0 \leq p_a \leq 15$

As Poincaré himself taught us [39], the patterns of Poincaré sections such as these are extremely involved. This can be appreciated in Fig. (3) [19], where a blow up of one of the “islands” created by the near resonance at the lower centre of Fig. (2) is displayed; the sector $.95 \leq a \leq 1.2$, $0 \leq p_a \leq .4$ is shown. We can see here the Poincaré section of orbits close to a separatrix; a stochastic layer must have formed around the separatrix itself, but it remains confined between surviving KAM tori, and thus chaotic motion is unable to spread. Of course, all this only confirms our previous analysis, but now with the added confidence of knowing that our numerical simulations are nonperturbative in the mass of the field.

Although the results above show convincingly that our model is chaotic, or at least non integrable, from a mathematical point of view, there are at least two objections

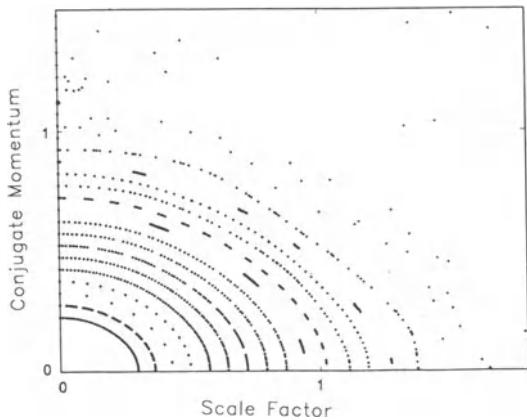


Figure 2. Poincaré sections on the $\phi = 0$ plane for $m = 0.8$. Initially $a = \dot{\phi} = 0$, $0.225 \leq p_\phi \leq 0.975$. Only the first quadrant is shown: $0 \leq a \leq 1.8$, $0 \leq p_a \leq 1.5$

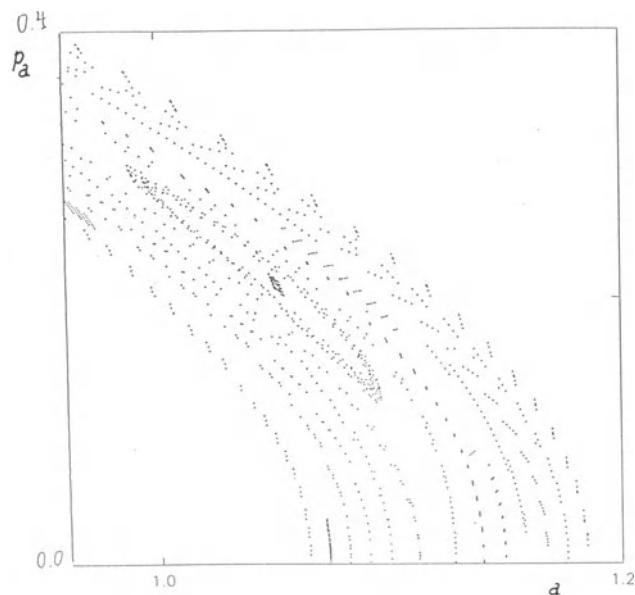


Figure 3. Poincaré sections on the $\phi = 0$ plane for $m = 0.8$. A blow up of the lower centre of the previous figure is shown; here, $.95 \leq a \leq 1.2$, $0 \leq p_a \leq .4$.

to be countered before the physical meaningfulness of chaos may be considered as established. The first criticism is simply that, since chaos arises at energy densities of order one in natural units, that is, near Planck's scale, it is improper not to consider quantum effects on the evolution of the Universe. We shall return to a discussion of this point later on, but by now we shall pass on to the second objection. This criticism states that, since the Melnikov method addresses a perturbation acting on an infinite time, that is, over an infinite number of cosmic cycles, it is irrelevant to the experiences of a physical observer confined to a single cosmic cycle.

The underlying source of this criticism is that, chaotic Hamiltonian systems being still deterministic, any chaotic evolution is totally regular over short enough time scales; what distinguishes chaotic from regular motion is the rate at which the accuracy in our knowledge of the orbits is dissipated by the evolution. Thus, if to the mathematician the relevant question concerns the infinite time behavior of the system, to the physicist the relevant question is whether substantial accuracy loss accrues on the time scales of observation, given typical observational accuracies.

For a system executing compact motion, the proper measure of accuracy loss is given by the Lyapunov exponents[40]. Indeed, the mean values of the positive Lyapunov exponents are related to the Kolmogorov entropy of the system[41]. Their inverses are the times scales over which nearby orbits diverge appreciably. In the present context, then, we may say that the cosmological relevance of chaos hinges on how the Lyapunov time compares with the mean lifetime of the Universe; if the former is much smaller than the latter, then chaos shall be relevant even for observers confined to single cosmic cycles. Moreover, a conservative Hamiltonian system with two degrees of freedom can have at most one positive Lyapunov exponent, so there is no ambiguity with respect to the scales concerned.

Unfortunately, it is extremely hard to evaluate Lyapunov exponents numerically. Instead, 'local Lyapunov exponents' are introduced, which average the eigenvalues of the linearized evolution operator around a solution, at several points along the orbit [42]. These local Lyapunov exponents are generally time - dependent; in actual computations, they display a strong transient behavior until enough data are accumulated. In very long runs, moreover, they become unreliable because of cumulative numerical errors. In other words, they are useful only when a "plateau" stands out clearly enough; in this case, the value at the plateau approximates the true Lyapunov exponent.

As shown by our analysis in the previous subsection, in the "strong chaos" region $m^2 K, m^2 j \gg 1$, the period of oscillation of the radius of the Universe becomes very well defined, $T \sim 4\pi$. Since this period actually counts an Universe and an "anti Universe" with $a \leq 0$, the time it takes for the Universe to reach maximum expansion after the Big Bang is of order π . Thus substantial accuracy loss follows for Lyapunov exponents greater than about a third. Now, simulations yield well defined Lyapunov exponents of the order of one half already for $m^2 j \sim 2$, very close to the "border of chaos" $m^2 j \sim 1/3$. Thus we conclude that, barring quantum phenomena, there is ample room for actual observations of chaos within a single Universe [19].

In order to obtain a more intuitive idea of how chaos might influence a concrete cosmology, we have performed the following "thought experiment". We considered an Universe starting out of a Big Bang ($a = 0$) with considerable momentum ($p_a = -5$). This yields $m^2 j \sim 12.5$, deep into the stochastic sea (for simplicity, we further take $m = 1$). We run some 70 simulations, allowing the field to take equally spaced initial values from 0 up to 3.55, and recorded the value of the field immediately before the

Big Crunch. According to our results, the values at recollapse are quite uncorrelated to those at the Big Bang; in particular, the correlation coefficient among the two series is of order 0.01, low enough to be seen as a mark of stochastic independence. This result underlies how tough it would be to make a precise prediction of the behavior of the Universe close to the Big Crunch, from data near the Big Bang [19].

The results of our experiment are remarkable in one other sense. Because our procedure effectively cuts out a small window around each singularity, it shows that the loss of information occurs, so to speak, “uniformly”; it is not tied up to the behavior near $a = 0$. This marks a stark contrast with the vacuum Bianchi IX models, where just the opposite is true. This is specially satisfying, since it is close to the singularities that the neglect of quantum phenomena is particularly worrisome.

We shall return to a discussion of the cosmological significance of chaos in our final remarks. But first, let us address a more concrete question, namely, how robust are the results shown against perturbations of the model. To seek an answer, we must extend our methods to cover non - Hamiltonian, as well as Hamiltonian, cosmological models. The discussion of this issue will be the subject matter of the remaining of this paper.

3.4. More General Hamiltonian Chaotic Cosmologies

While appealing due to its sheer simplicity, the model presented above is still too special to be seen as proof of the prevalence of chaos in realistic cosmological models. At least, we should be able to show that chaos is robust against small perturbations of the model, before any claim of generality could be advanced. In this section, we shall discuss briefly two possible generalizations of the model, namely, adding a cosmological constant, and considering non conformal ($\xi \neq 1/6$) couplings. These generalizations will allow us to remain within the framework of classical Hamiltonian cosmology. A third possible direction in which our model may be generalized is to add further kinds of matter; this will force us to consider non Hamiltonian models, and will be the subject matter of next Section.

3.4.1. Adding a cosmological constant. Long considered Einstein’s worst mistake, the inclusion of a cosmological constant in Einstein’s equations became again fashionable when it was realized that, in a theory including quantum matter, its vanishing posed a rather nontrivial puzzle. The development of Inflationary cosmologies further legitimized the cosmological constant, the only remaining question being whether it survived the inflationary era, or was totally dissipated into heat during reheating [43].

From our point of view, adding a cosmological constant is the simplest modification to be made on our model. It amounts to add a new term to the Hamiltonian of Eq. (22), which now reads

$$H = (1/2)[(p_\phi^2 + \phi^2) - (p_a^2 + a^2) + m^2 a^2 \phi^2 + 2\Lambda a^4]. \quad (36)$$

Because now the theory is non linear for $m^2 = 0$, it makes sense to simply take the mass as perturbative parameter. We must then begin by studying the solutions for $m^2 = 0$.

It is obvious that the system is integrable in this limit. The field is just an harmonic oscillator, so we simply parametrize $\phi = \sqrt{2j} \sin \varphi$, $p_\phi = \sqrt{2j} \cos \varphi$. We must further distinguish between “hot” Universes with $16j\Lambda > 1$, and “cold” ones with

$16j\Lambda < 1$. The former expand forever, while the latter either recollapse, or never hit $a = 0$. Clearly, for “hot” Universes, there is no question of chaotic behavior, as they will approach asymptotically a DeSitter spacetime [44]. For “cold” recollapsing Universes, on the other hand, the influence of the cosmological term is small, and the analysis above will still apply.

For the critical value $16j\Lambda = 1$, we find two fixed points, representing an Einstein or an “anti” Einstein Universe ($2a\sqrt{\Lambda} = \pm 1$, respectively). They are joined by an “instanton”-like solution $a(\eta) = (1/2\sqrt{\Lambda}) \tanh(\eta/\sqrt{2})$. This separatrix is destroyed by the perturbation $m^2 ja^2 \sin^2 \varphi$; concretely, the Melnikov integral yields (cfr. Eq. (7)) $\Delta E \sim (m^2/8\Lambda^2)(\sqrt{2}\pi/\sinh\sqrt{2}\pi)\sin\varphi_0$.

Therefore, these models either expand forever, approaching DeSitter behavior, or else are chaotic. Chaos follows essentially the patterns already discussed, and is strongest close to the separatrix. An intriguing possibility is that chaotic diffusion in phase space [5, 7, 45] could allow “cold” Universes to become “hot”, leaking through the former separatrix. This would not happen in the present model, as diffusion is suppressed in low dimensions, but becomes possible if further matter fields are included. If confirmed, this phenomenon would be a classical analog of the known enhancement of Inflation due to semiclassical particle creation[46].

3.4.2. Non conformal coupling . The extension of our results to non conformal couplings would be highly desirable, as this would greatly enhance its cosmological appeal. In particular, most Inflationary models assume a minimally coupled ($\xi = 0$) inflaton, so this would be the most interesting case to study.

Indeed, the minimally coupled case has been recently analyzed by Lyons [47]. Writing the field in terms of suitable adiabatic amplitude and phase, he succeeded in reducing the dynamics to a perturbation acting on an integrable Hamiltonian. The unperturbed Hamiltonian describes dust dominated FRW cosmologies, and is easily solved. The perturbation, however, diverges at the cosmic singularities [48]. Because of this divergence, it is not possible to use the fiction of “cosmic cycles” to couch the problem in a manner suitable to the application of Melnikov’s method.

The similarities between the minimal and conformal models unveiled by Lyons make it plausible that, away from the singularities, both models will have essentially the same behavior, including large local Lyapunov exponents and substantial mixing. However, as the minimal model lies outside the scope of Melnikov’s method, we shall not discuss it further here.

4. HOMOCLINIC CHAOS IN NON HAMILTONIAN COSMOLOGY

The third and final modification of our chaotic FRW model to be discussed in these lectures concerns the possibility of including new kinds of matter. As already mentioned, the inclusion of higher spin fields should not modify our conclusions, except that presumably the model will become even more chaotic as new phenomena, such as Arnold’s diffusion, come into play. Much more interesting is to consider the inclusion of dissipative fluid matter, as this will force us out of the realm of Hamiltonian cosmology. Dissipation could arise out of ordinary phenomena, such as bulk viscosity arising from the mixing of two perfect fluids of different masses [49], or else could be thought of as a device to include semiclassical effects such as particle creation into a classical model [15].

In general, dissipation will add terms related to heat flow, shear and bulk viscosity to the matter energy momentum tensor. These terms will be related to the gradients of the relevant thermodynamic variables (temperature, energy and particle number densities, and the velocity of the fluid). If we assume linear relations between these sets of variables (stresses and gradients), we arrive to the well known Eckart and Landau - Lifshitz formalisms, which provide a simple generalization of the corresponding non relativistic theory[50]. Unfortunately, these “linear” theories also posses severe stability and causality problems. There is, however, a framework to systematically improve on the linear theories, provided by more general approaches to non equilibrium thermodynamics, such as Extended Thermodynamics [51]. At present, though, the question of whether a satisfactory, logically closed, relativistic theory of real fluids will ever be formulated, is regarded as undecided by many authors[52].

In the case of FRW cosmology, heat flux and shear are impossible, given the symmetries of the problem, and the only possible viscous effect is the inclusion of a dissipative pressure $\Pi = \zeta H$, where $H = a'/a^2$ denotes Hubble's constant. There is at present no systematic derivation of the dependence on ζ on other thermodynamic variables in realistic settings [53]; most phenomenological approaches assume $\zeta = \alpha\tau^m\rho^n$, where α is a constant, τ is the mean free time for the particles of the fluid, and ρ is the energy density. For simplicity, we shall make the rather unrealistic assumption $m = n = 0$; thus, for us, ζ will be simply a constant.

To define the model we must also give the equation of state of the fluid. Again for simplicity, and to avoid including new scales in the theory, we shall assume that the fluid behaves as conformally invariant radiation, $p = \rho/3$. The Raychauduri equation now becomes

$$a'' - \zeta aa' + a - m^2\phi^2a = 0. \quad (37)$$

The Klein - Gordon equation is unchanged, and the 00 Einstein's equation becomes

$$(1/2)[(\phi'^2 + \phi^2) - (a'^2 + a^2) + m^2a^2\phi^2] = -\rho a^4. \quad (38)$$

Observe that ρ diverges at the cosmic singularities. However, as this divergence does not affect the dynamical equations, it does not hinder the analytical continuation of the metric beyond these points.

Because our fluid behaves in equilibrium as radiation, it is possible to associate with it an entropy density $s \sim (4/3)\rho^{3/4}$. The Second Law takes the form $S' = [a^3 s]' = 3\zeta a'^2 a/[a^4 \rho]^{1/4}$. Observe that $S' \leq 0$ when $a \leq 0$, as it should, in order that $dS/dt = S'/a$ be positive, t being the proper time of a co-moving observer.

Even this simple model is already too complex to be studied in full generality, and we must resort to perturbative approaches. Two simple limiting cases stand out, when $\zeta \rightarrow 0$, m^2 fixed, and the opposite one, $m^2 \rightarrow 0$, ζ constant. In spite of appearances, in both cases the unperturbed system may be shown to be Hamiltonian; perturbations, though, will be non Hamiltonian, which will force us to generalize somewhat the framework presented in Section (2.4). We shall consider each of these limiting cases in turn, beginning with the simplest, $\zeta \rightarrow 0$ one.

4.1. Vanishing ζ limit

In the present context, this case is the simplest, because it allows us to use in full the analysis made in the previous Section. If we describe the evolution in terms of the

canonical variables (a, p_a, ϕ, p_ϕ) , we find the same equations of motion, except now $p'_a = -(\partial H/\partial a) + 3\zeta R$, where $R = ap_a$. It follows that, in any set of canonical coordinates q_i , the equations of motion take the form $q'_i = \{H, q_i\} + \delta q'_i$, where $\delta q_i = 3\zeta(\partial q_i/\partial p_a)R$. Here, the derivative is taken with (a, ϕ, p_ϕ) held constant.

In particular, bulk viscosity modifies the evolution of the momentum $-L$, which, as we already saw, plays the role of reduced Hamiltonian for the resonating variables (j, Θ) . Actual computation yields for $-L'$ an oscillatory part plus an steady drift $\delta(-L') \sim 3\zeta m j_n K_n$ for a large n resonance. Because of this drift in the reduced Hamiltonian, the recurrence of the homoclinic motion is destroyed, and chaos erased, as soon as ζ takes a non vanishing value[12].

As a matter of fact, we find here the already mentioned divergence between the mathematical and physical views of chaos. The above conclusion follows from allowing an infinite span to the angle variable $\tau = -\alpha$, that is, going through an infinite number of cosmic cycles. A more physical view would be to compare the cumulative drift $\Delta(-L)$ over a single cosmic cycle, with a typical energy, e. g., the width $\delta E \sim (1/m^2)$ of the Hamiltonian stochastic layers. Now, since the period of oscillation is nearly constant, we find $\Delta(-L)/\delta E \sim n^3(\zeta/m)$. So, as $\zeta \rightarrow 0$, we still can find interesting dynamical behavior in a bounded region in phase space.

The divergence of the Melnikov integral, in any case, shows that the limit $\zeta \rightarrow 0$, m^2 fixed, is not analytical. Dissipation allows the amplitude of the a oscillations to grow, until eventually the $\zeta aa'$ term in the equations of motion is no longer small. This makes all the more important to study the opposite limit, ζ fixed, $m^2 \rightarrow 0$, to which we proceed.

4.2. Vanishing m^2 limit

In this subsection, we shall study the dynamical system defined by the Klein - Gordon equation Eq. (20), and the dissipative Raychauduri equation Eq. (37), in the limit when $m^2 \rightarrow 0$. Let us then start with an analysis of the massless case.

If $m^2 = 0$, then the field decouples from the radius of the Universe. It is clear that the field evolution is Hamiltonian. Parametrizing the field and its conjugated momentum in terms of action angle variables (j, φ) , as in Section (3.4.1), the hamiltonian is simply $H_f = j$.

As for the radius of the Universe a , we observe that the main effect of the bulk viscosity term is to provide a negative contribution to the pressure. Thus, we could expect to find the same kinds of behavior than in presence of a cosmological constant. Indeed, an Universe emerging from $a = 0$ with $a' > 1/\zeta$ will expand forever, approaching De Sitter behavior; Universes with $a' < 1/\zeta$ eventually recollapse. The role of “separatrix” is played by the particular solution $a = \eta/\zeta$. Bulk viscosity driven Inflation has received some attention in the literature [54]; here, we shall concern ourselves only with recollapsing Universes.

In spite of appearances, massless Eq. (37) may be derived from the Lagrangian

$$\mathcal{L}_a = \left(\frac{-1}{\zeta^2}\right)\{[1 - \zeta a'] \ln[1 - \zeta a'] + \zeta a'\} + \left(\frac{1}{2}\right)a^2. \quad (39)$$

Thus we can associate to a the conjugated momentum

$$\pi_a = \left(\frac{1}{\zeta}\right)\ln[1 - \zeta a'], \quad (40)$$

leading to the Hamiltonian

$$\mathcal{H}_a = \left(\frac{-1}{\zeta^2}\right)\{e^{\zeta\pi_a} - 1 - \zeta\pi_a\} - \left(\frac{1}{2}\right)a^2. \quad (41)$$

Of course, in the $\zeta \rightarrow 0$ limit we recover the usual Hamiltonian $(-1/2)(\pi_a^2 + a^2)$. Also, in this limit, $\pi_a = p_a = -a'$. The Hamiltonian Eq. (41) generates bounded motions; this is most clearly seen if we regard π_a as the canonical variable, and $-a$ as the conjugated momentum, which reduces the dynamics to a particle moving into a confining potential well.

Being conserved, \mathcal{H}_a is integrable. We may introduce action - angle variables (K, α) , and Fourier expand (a, π_a) . The Fourier expansions are related by the Hamiltonian equation $\pi'_a = a$, since, in turn, $\pi'_a = (\partial\pi_a/\partial\alpha)\alpha'$, and $\alpha' = -\Omega(K)$. Thus, if

$$a = \sum_{n=1}^{\infty} a_n(K) \sin n\alpha \quad (42)$$

then

$$\pi_a = \pi_{a0} + \sum_{n=1}^{\infty} \frac{a_n(K)}{n\Omega(K)} \cos n\alpha \quad (43)$$

To find the equations of motion in the presence of the mass terms, we shall assume that the “definition” Eq. (40), and its companion $p_\phi = \phi'$, still hold. Then the dynamical equations for π_a and p_ϕ become

$$\pi'_a = a - m^2 \phi^2 a e^{-\zeta\pi_a} \quad (44)$$

$$p'_\phi = -[1 + m^2 a^2] \phi \quad (45)$$

In terms of the massless action - angle variables (K, α, j, φ) , the equations become

$$\varphi' = 1 + m^2 a^2(K, \alpha) \sin^2 \varphi \quad (46)$$

$$j' = -m^2 j a^2(K, \alpha) \sin 2\varphi \quad (47)$$

$$\alpha' = -\Omega(K) + m^2 j \left(\frac{\partial a^2}{\partial K}\right)(K, \alpha) (\sin^2 \varphi) e^{-\zeta\pi_a(K, \alpha)} \quad (48)$$

$$K' = -m^2 j \left(\frac{\partial a^2}{\partial \alpha}\right)(K, \alpha) (\sin^2 \varphi) e^{-\zeta\pi_a(K, \alpha)} \quad (49)$$

This system of equations is not quite Hamiltonian; to disentangle its Hamiltonian and dissipative parts it is convenient to introduce a new variable

$$\kappa = K + m^2 j \sin^2 \varphi \left(\frac{d\Omega}{dK}\right)^{-1} \left(\frac{\partial a^2}{\partial K}\right)[1 - e^{-\zeta\pi_a}] + O(m^4) \quad (50)$$

Retaining only terms of first order in m^2 , we observe that Eqs. (46) and (47) retain their form, while Eqs. (48) and (49) become, respectively

$$\alpha' = -\Omega(\kappa) + m^2 j \left(\frac{\partial a^2}{\partial K}\right)(\kappa, \alpha) (\sin^2 \varphi) \quad (51)$$

$$\kappa' = -m^2 j \left(\frac{\partial a^2}{\partial \alpha} \right) (\kappa, \alpha) (\sin^2 \varphi) - \delta \kappa' \quad (52)$$

where

$$\delta \kappa' = m^2 j \left\{ \sin^2 \varphi \left[\frac{\partial}{\partial \alpha} \left[\left(\frac{\Omega}{\Omega_K} \left(\frac{\partial a^2}{\partial K} \right) - a^2 \right) [1 - e^{-\zeta \pi_a}] \right] + \frac{\zeta a^3}{\Omega} e^{-\zeta \pi_a} \right] - \frac{\sin 2\varphi}{\Omega_K} \left(\frac{\partial a^2}{\partial K} \right) [1 - e^{-\zeta \pi_a}] \right\} \quad (53)$$

Clearly, we are dealing with a Hamiltonian system, with Hamiltonian

$$H = \mathcal{H}_a(\kappa) + j + m^2 j a^2 (\kappa, \alpha) \sin^2 \varphi \quad (54)$$

subject to the non - Hamiltonian perturbation Eq. (53). In the $\zeta \rightarrow 0$ limit, of course, we recover the Hamiltonian from the previous Section.

Taking the first two terms in the Hamiltonian as the unperturbed one, we find resonances will occur when $2 - n\Omega = 0$. This equation must have solutions, since we know that, as $\kappa \rightarrow \infty$, the evolution of the Universe approaches the separatrix $a \sim \eta/\zeta$, which corresponds to $\Omega = 0$ (indeed, it can be shown that $\Omega \sim \kappa^{-1/3}$ in the relevant range). We may analyze the behavior close to a resonance, following the procedure established in Sections (2.4) and (3.2); when we include the non - Hamiltonian term in the evolution, we find that the reduced energy of the resonant degree of freedom is no longer conserved. However, now the situation is totally different from that in Section (4.1), because $\delta \kappa'$, as given by Eq. (53), averages to zero over periodic motions. This is obvious for those terms which are actually α or ϕ derivatives, and follows from Eq. (42) for the remainder. Therefore, in this case, dissipation does not preclude recurrent motions, and the by now standard analysis of the dynamics close to a resonance will indicate the formation of stochastic layers.

It should be stressed that, while in the Hamiltonian case we only found chaos above a certain threshold ($m^2 j \geq (1/3)$), in the dissipative case the stochastic layers seem to extend up to $j \sim 0$. In this sense, the dissipative model is more chaotic than its conservative counterpart, as could be expected, given the added non linearity viscosity brings to the Raychauduri equation. However, as we already remarked, these perturbative results should be regarded as mere hints of interesting dynamical behavior, pending confirmation by more reliable, non perturbative tests.

An aspect where our results in this section resemble those of the Hamiltonian case, is that in both chaos may seem to prevail in a region of phase space overlapping the quantum era. Let us conclude this Section with a brief discussion of why this was to be expected, and what could be learned from it.

4.3. Classical Chaos and Quantum Cosmology

It should be clear that in any realistic cosmological model there are many sources of chaotic behavior of hydrodynamical origin, which are simply the transposition to the relativistic regime of phenomena already described in Navier - Stokes theory [55]. Similarly, , any self gravitating system involving several bodies, being already chaotic at the Newtonian level, will presumably continue being so in relativistic theory [56]. Our interest in these lectures is focused not in these phenomena, but on the contrary on chaotic behavior stemming from the dynamics of the gravitational field itself. This declared interest pushes us to consider the realms where non linear gravitational effects are strongest, namely, the very early Universe and the neighborhood of collapsed

objects. As a matter of fact, these are also the regimes where quantum gravitational effects become noticeable. The coexistence of quantum and chaotic behavior suggests the possibility of an interesting interaction between these two aspects of gravitational physics.

As it turns out, the relationship between quantum, semiclassical and classically chaotic behavior is complex and proceeds at several levels. Most obvious, semiclassical and quantum effects should be incorporated in classical models of the very early Universe as phenomenological terms. Let us comment briefly on the different ways the leading features of quantum behavior may be introduced into otherwise classical models.

The simplest way in which semiclassical effects influence classical cosmology, is through the fact that the former determine the observable values of the different coupling constants (including Newton's and the Cosmological one) for the latter [57]. Moreover, semiclassical effects induce new terms in the effective gravitational action, beyond the Einstein - Hilbert functional. Typically, these terms involve higher derivatives of the canonical variables, and are associated to severe stability and causality problems. However, it has been shown that at least some of the solutions of the extended theories are physical, and represent genuine predictions of semiclassical gravity [58].

Notwithstanding, enlarging the Einstein - Hilbert action to an effective, even non local, action functional does not capture the full complexity of semiclassical dynamics. A semiclassical model consists of classical fields interacting with the infinite degrees of freedom of their quantum fluctuations. In ordinary settings, the detailed state of these fluctuations is not observable, and so the evolution of the classical fields appears to the observer as inherently dissipative [59]. In semiclassical cosmology, the leading dissipative mechanism is particle creation out of the gravitational field[60]. Particle creation has a smoothing effect on the Universe, favoring the dissipation of inhomogeneities and anisotropies, and the onset of Inflationary periods [61].

Perhaps less expected, semiclassical evolution is also explicitly stochastic [62, 63]. Indeed, the emergence of a classical Universe from the quantum era concerns the fact that our knowledge of the Universe is restricted to certain coarse grained aspects of it [64]; the neglected sector, however, reacts back on the observable one, and induces fluctuations in the macroscopic variables. Through a well understood procedure, these fluctuations may be included into the semiclassical equations of motion as so many classical stochastic sources, their statistical properties being determined by the underlying quantum state of the Universe. It is conjectured that fluctuations of this kind may have been the origin of density fluctuations in the Inflation era, which eventually seeded the present day large scale structures [65, 66]. As a matter of fact, simple Inflationary models predict a Harrison - Zel'dovich spectrum of primordial fluctuations, in agreement with the present view of the fluctuations in the Cosmic Microwave Background [30].

It should be clear from the above that semiclassical cosmological models are a fertile ground for the applications of dynamical systems theory. These models are highly non linear, to begin with; as we have seen, dissipation is no obstacle to chaos, and may even enhance it. Finally, it is clear that the strong instability of typical orbits in a chaotic system may amplify the primordial fluctuations of quantum origin in unexpected ways, and change our view of pattern formation in the early Universe [67].

Now, the flip side of the coin is, quantum cosmology has many things to learn from chaos. Being a theory besieged by interpretational difficulties, the view is often expressed that, in any case, it does not matter to be able to interpret a theory before

there are any physicists around to perform experiments on it. In other words, what should matter of Quantum Cosmology is that in some regime it should predict the probabilities of different initial conditions for classical cosmic evolution, physicists being macroscopic objects that can only exist in a mostly classical environment.

It is generally believed that Quantum Cosmology performs this task through the fact that, for large enough Universes, the Wave Function adopts a WKB form [68]. This WKB Wave Function, or the Wigner Function constructed from it, is associated to a bundle of classical trajectories, which are, essentially, the possible classical histories of the Universe [69].

This simple scenario has many shortcomings, which have been pointed out over time. Basically, even in the WKB regime a quantum state retains its coherence properties, and therefore predicts a number of phenomena, namely, interference effects, which have no classical interpretation. Therefore, the assimilation of the WKB state to a set of classical histories presumes that interference between different histories has been suppressed (“decoherence”). Now decoherence involves the restriction of our description of the Universe to only certain “consistent” sets of coarse grained histories; as we have already remarked, in principle, the evolution of a coarse grained history is not deterministic, and thus classicality seems to hinge on the balance between decoherence and fluctuations [70, 71].

Nevertheless, it is when chaos is brought into the picture that this view of classical cosmology arising out of the quantum era becomes outright problematic. As has been demonstrated through the study of Quantum Maps, when the underlying classical dynamics is ergodic, the correspondence between quantum and classical evolution in the WKB regime is only approximate, and breaks down at large enough times [72]. As a matter of fact, Quantum Maps may be seen as approximating the behavior of a particle moving close to a resonant homoclinic loop [7], and so, the conclusions derived from them can be easily translated to systems displaying Homoclinic Chaos. Therefore, it is to be expected that a clear statement of the semiclassical limit of Quantum Cosmology, embracing such models as described in the body of these lectures, will require a substantial revision of the usual “WKB interpretation”.

Beyond what Quantum Cosmology might teach Chaos, or vice versa, lie a number of issues which are problematic to both fields alike. They concern the issue of time (which surfaces in chaotic cosmology under the guise of when a Lyapunov exponent is a suitable characterization of chaos) and the Hamiltonian structure of the theory. This latter subject is particularly pressing if we are ever going to move beyond Mini Super Space and linearized approximations, into the Universe of the full Einstein’s equations.

In this connection, the classical gravitational chaos community would be wise to keep an eye on the development of the “new canonical (“Ashtekar”) variables” approach to quantum gravity[73]. In this approach, the structure of the constraints becomes substantially simpler than in the more common ADM formalism, although this gain is partially offset by the presence of new consistency conditions (“reality conditions”) to be imposed on physically meaningful solutions. While to date little progress has been achieved beyond Bianchi and lower dimensional models, it is a safe bet that approaches of this kind will be essential in understanding the real incidence of chaos in the full Einstein system [74, 75].

As can be seen from the above, the fact that the region where gravitational chaotic behavior becomes significant overlaps the quantum era, far from being a limitation on

the usefulness of the theory, should be regarded as one of the most powerful motivations for its pursuit.

5. FINAL REMARKS

In the lectures above, we have reviewed Homoclinic Chaos, we have shown how it frequently arises through internal resonances in Hamiltonian systems of several degrees of freedom, and we have described in some detail two simple cosmological models, one Hamiltonian, the other dissipative, where it occurs. Our results suggest that Homoclinic Chaos is widespread in relativistic Cosmology, and point the way to some simple criteria for its detection. We have endorsed throughout a mix of analytic and numerical methods as the soundest strategy in this kind of research. Also we have argued in favor of the use of gauge invariant criteria of chaos, whenever possible, although non gauge invariant ones are acceptable in a well defined observational context.

The proper recognition of the richness of behavior associated to classical models of the early Universe may greatly affect our understanding of such cosmic features as the distribution of large scale structures, and the isotropy of the Cosmic Microwave Background. As we discussed in the previous Section, these studies may affect our understanding of the very early Universe, even beyond the validity of classical cosmology itself. Moreover, realistic cosmological models will include many fold chaotic phenomena associated to hydrodynamical turbulence, over and above the purely geometrodynamical chaos described here.

It is likely, however, that our understanding of the structure of Einstein's equations is still too rudimentary to support realistic model building. In particular, we still have incomplete knowledge of the typical behavior of solutions of Einstein's equations far from exceptional symmetries, or the Brill wave regime. To really investigate such issues would force us to confront chaos in systems of infinite number of degrees of freedom, with the attending formidable challenge to both our analytical and numerical tools [76].

Beyond the particular new phenomena that may be associated to Chaos in self gravitating systems, notwithstanding, it remains the fact that methods inspired by Dynamical Systems Theory, such as those used in these lectures, remain a powerful guidance in the analysis of complex systems, allowing the identification of the essential elements of the dynamics, and leading to both qualitative and quantitative predictions concerning typical behavior. It is in this capacity, rather than from a narrow focus on Chaos as such, that we may expect the most important contributions from our subject to the traditional concerns of General Relativity and Cosmology.

We intend to continue our research on this rewarding field.

ACKNOWLEDGEMENTS

I would like to thank the organizers for their most kind invitation to join the workshop, and for their hard work and enthusiasm which made it such a remarkable success. I would also like to acknowledge the many discussions with my colleagues L. Bombelli, C. El Hasi and R. Tavakol, with whom the original work on which these lectures are based was developed. My special regards to C. El Hasi, for his help in setting the figures and checking the calculations for this manuscript.

This work was partially supported by the Directorate General for Science Research and Development of the Comission of the European Communities, under Contract N° C11-0540-C, and by CONICET, UBA and Fundación Antorchas.

REFERENCES

- [1] Guckenheimer J. and Holmes P., 1983, *Non-Linear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Berlin: Springer-Verlag).
- [2] Wiggins S., 1988, *Global Bifurcations and Chaos* (Heidelberg: Springer-Verlag).
- [3] Holmes, P., 1990, Poincaré, Celestial Mechanics, Dynamical Systems Theory and "Chaos" *Phys. Rep.* **193**, 137.
- [4] Matinyan, S. G., Savvidi, G. K. and Ter - Arutyunyan - Savvidi, N. G., 1981, Classical Yang - Mills Mechanics, Nonlinear Color Oscillations *Zh. Eksp. Teor. Fiz.* **80**, 830 (Engl. trans. *Sov. Phys. JETP* **53**, 421).
Kawabe, T. and Ohta, S., 1991, Order - to - Chaos Transition in SU(2) Yang - Mills Theory *Phys. Rev. D* **44**, 1274.
Bambah, B. A., Lakshmibala, Mukku, C. and Sriram, M. S., 1993 Chaotic Behavior in Chern - Simons - Higgs Systems *Phys. Rev. D* **47**, 4677.
- [5] Chirikov B. V., 1979, A universal instability of many dimensional oscillator systems *Phys. Rep.* **52**, 263.
- [6] Reichl, L. E., and Zheng, W. M., 1987, Non Linear Resonance and Chaos in Conservative Systems, in *Directions in Chaos* ed. Hao Bai-Lin (Singapore, World Scientific).
- [7] Zaslavsky, G. M., Sagdeev, R. Z., Usikov, D. A. and Chernikov, A. A., 1991, *Weak Chaos and Quasi Regular Patterns* (Cambridge: Cambridge University Press).
- [8] Ornstein, D., 1974, *Ergodic Theory, Randomness, and Dynamical Systems* (New Haven, Yale University Press).
- [9] Shields, P., 1974, *The Theory of Bernoulli Shifts* (Chicago: University of Chicago Press).
- [10] Arnold, V. I., 1978, *Mathematical Methods of Classical Mechanics* (Berlin: Springer-Verlag) (Second edition, 1989).
Arnold, V. I., Kozlov, V. V. and Neishtadt, A. I., 1988, *Mathematical Aspects of Classical and Celestial Mechanics*, Dynamical Systems III, Encyclopaedia of Mathematical Sciences (Heidelberg: Springer-Verlag)
- [11] Holmes, P. J. and Marsden, J. E., 1982, Melnikov's method and Arnold diffusion for perturbations of integrable Hamiltonian systems *J. Math. Phys.* **23** 669-75.
- [12] Holmes, P. J. and Marsden, J. E., 1982, Horseshoes in perturbations of Hamiltonian systems with two degrees of freedom *Commun. Math. Phys.* **82** 523-44.
- [13] Misner, C., 1972, Minisuperspace, in *Magic without Magic*, ed. J. Klauder (Freeman, San Francisco), p 441.
- [14] Palais, R. S., 1979 The Principle of Symmetric Criticality *Commun. Math. Phys.* **69**, 19.
- [15] Hu, B. L., 1982 *Phys. Lett.* **90A**, 375.
- [16] Arnold, V. I. and Avez, A., 1968, *Ergodic Problems of Classical Mechanics* (New York: Benjamin)

- [17] Regge, T. and Wheeler, J. A., 1957, Stability of a Schwarzschild singularity *Phys. Rev.* **108** 1063-9
Vishveshwara, C. V., 1970, Stability of the Schwarzschild metric *Phys. Rev. D* **1** 2870-9
Moncrief, V., 1974, Gravitational perturbations of spherically symmetric systems. I. The exterior problem *Ann. Phys. (N.Y.)* **88** 323-42
Chandrasekhar, S., 1983, *The Mathematical Theory of Black Holes* (Oxford: Clarendon Press).
- [18] Bombelli, L. and Calzetta, E., 1992, Chaos around a Black Hole *Class. Quantum Grav.* **9**, 2573.
- [19] Calzetta, E. and El Hasi, C., 1993 Chaotic Friedmann-Robertson-Walker Cosmology *Class. Quantum Grav.* **10**, 1825.
- [20] Calzetta E., El Hasi, C. and Tavakol, R, 1993, *to appear*
- [21] Ozorio de Almeida, A. M., 1988, *Hamiltonian Systems, Chaos and Quantization* (Cambridge: Cambridge University Press).
- [22] Prigogine, I. and Elsken, Y., 1987, Irreversibility, Stochasticity and Non Locality in Classical Dynamics, in *Quantum Implications* ed. B. J. Hiley and F. D. Peat (London: Routledge)
- [23] Sinai, Ya. G., 1970 *Theory of Dynamical Systems*, Lecture Notes Series 23 (Warsaw University)
Cornfeld, I. P., Sinai, Ya. G. and Fomin, S. V., 1982, *Ergodic Theory* (Heidelberg: Springer-Verlag).
- [24] Khinchin, A., 1957, *Mathematical Foundations of Information Theory* (New York: Dover)
- [25] Misner, C., Thorne, K. and Wheeler, A., 1972, *Gravitation* (San Francisco, Freeman).
- [26] Baierlein, R. F., Sharp, D. H. and Wheeler, J. A., 1962, Three Dimensional geometry as Carrier of Information about Time, *Phys. Rev.* **126**, 1864.
- [27] Wheeler, J. A., 1968, Super Space and the Nature of Quantum Geometrodynamics, in *Batelle Rencontres* ed. C. DeWitt and J. A. Wheeler (New York, Benjamin).
- [28] Ryan, M., 1972, *Hamiltonian Cosmology* (Berlin, Springer - Verlag)
Ryan, M. and Shepley, L., 1975, *Relativistic Homogeneous Cosmology* (Princeton, Princeton University Press).
- [29] MacCallum, M. A. H., 1979, Anisotropic and Inhomogeneous Relativistic Cosmologies in *General Relativity* ed. S. W. Hawking and W. Israel (Cambridge, Cambridge University Press) p. 533.
- [30] G. F. Smoot, 1993, COBE DMR Observations of the Early Universe, *Class. Quantum Grav.* **10** (1993).
- [31] Börner, G., 1988, *The Early Universe* (New York, Springer - Verlag) (2nd. Edition 1992).
- [32] Hawking, S. W., 1987, Quantum Cosmology, in *300 Years of Gravitation*, ed. S. W. Hawking and W. Israel (Cambridge, Cambridge University Press) p. 631.
- [33] Parker, L. and Ford, L. H., 1977 *Phys. Rev. D* **16**, 245
Parker, L. and Ford, L. H., 1977 *Phys. Rev. D* **16**, 1601.
- [34] Belinsky, V. A., Grishchuk, L. P., Khalatnikov, I. M. and Zel'dovich, Ya. B., 1985, Inflationary Stages in Cosmological Models with a Scalar Field, *Phys. Lett.* **155B**, 232-6
Belinsky, V. A., Grishchuk, L. P., Khalatnikov, I. M. and Zel'dovich, Ya. B., 1985 (same title) *Zh. Eksp. Teor. Fiz.* **89** 346-60 (Engl. Trans. Sov. Phys. *JETP* **62** 195-203)

- Belinsky, V. A., Grishchuk, L. P., Khalatnikov, I. M. and Zel'dovich, Ya. B., 1985 (same title) Proceedings of the Third Seminar on Quantum Gravity ed. M. A. Markov, V. A. Berezin and V. P. Frolov (Singapore, World Scientific) 566-90
- Gottlober, S., Muller, V. and Starobinsky, A., 1991, Analysis of Inflation Driven by a Scalar Field and a Curvature Squared Term *Phys. Rev. D* **43** 2510-20.
- [35] Futamase, T., Rothman, T. and Matzner, R., 1989, Behavior of Chaotic Inflation in Anisotropic Cosmologies with Nonminimal Coupling *Phys. Rev. D* **39** 405-11
- Maeda, K., Stein-Schabes, J. and Futamase, T. 1989, Inflation in a Renormalizable Cosmological Model and the Cosmic No Hair Conjecture *Phys. Rev. D* **39** 2848-53
- Amendola, L., Litterio, M. and Occhionero, F., 1990, The Phase Space View of Inflation (I) *Int. J. Mod. Phys. A* **5** 3861-86
- Demianski, M., 1991, Scalar Field, Nonminimal Coupling, and Cosmology *Phys. Rev. D* **44** 3136-46
- Demianski, M., de Ritis, R., Rubano, C. and Scudellaro, P. 1992 Scalar Fields and Anisotropy in Cosmological Models *Phys. Rev. D* **46** 1391-8.
- [36] Linde, A. D., 1982 *Phys. Lett.* **108B**, 389.
- [37] Courant, R. and Hilbert, D., 1953, *Methods of Mathematical Physics* (New York, Wiley) Vol I, p. 531.
- [38] Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T., 1985 *Numerical Recipes: The Art of Scientific Computing* (Cambridge, Cambridge University Press)
- [39] Poincaré, H., 1892, *Les Méthodes Nouvelles de la Mécanique Céleste* (Gauthier-Villars, Paris).
- [40] Lichtenberg, A. J. and Lieberman, M. A., 1992 *Regular and Chaotic Dynamics* (New York, Springer - Verlag).
- [41] Pesin, Ya. B., 1977 Characteristic Lyapunov Exponents and Smooth Ergodic Theory, *Uspekhi Mat. Nauk.* **32:4**, 55 (Engl. Trans. *Russian Math. Surveys* **32:4**, 55).
- [42] Wolf, A., Swift, J., Swinney, H. and Vastano, J., 1985, Determining Lyapunov Exponents from a Time Series *Physica* **16D**, 285-317.
- [43] Weinberg, S., 1989, The Cosmological Constant Problem *Rev. Mod. Phys.* **61**, 1.
- [44] Gibbons, G. and Hawking, S. W., 1977, *Phys. Rev. D* **15**, 2738.
- Starobinsky, A. A., 1983, *Piz'ma Zh. Eksp. Teor. Fiz.* **37**, 55 (Engl. Trans. *JETP Letters* **37**, 66).
- [45] Lieberman, M. A. and Tennyson, J. L., 1983, Chaotic motion along resonance layers in near integrable Hamiltonian systems with 3 or more degrees of freedom, *Long-Time Prediction in Dynamics* ed. C W Horton, L E Reichl and V G Szebehely (New York: John Wiley), p. 179.
- [46] Calzetta, E., 1991, Particle Creation, Inflation, and Cosmic Isotropy *Phys. Rev. D* **44**, 3043.
- [47] Lyons, G., *private communication*.
- [48] Laflamme, R. and Shellard, E. P. S., 1987, Quantum Cosmology and Recollapse, *Phys. Rev. D* **35**, 2315
- Hawking, S. W., Laflamme, R. and Lyons, G. W., 1993, Origin of Time Asymmetry *Phys. Rev. D* **47**, 5342.
- [49] Weinberg, S., 1971, Entropy generation and the Survival of Proto - Galaxies in an Expanding Universe, *Ap. J.* **168**, 175.
- [50] Weinberg, S., 1972 *Gravitation and Cosmology* (New York, John Wiley).

- [51] Jou, D., Casas - Vázquez, J. and Lebon, G., 1993, *Extended Irreversible Thermodynamics* (Berlin, Springer - Verlag).
- [52] Israel, W., 1988, Covariant Fluid Mechanics and Thermodynamics: an Introduction, in *Relativistic Fluid Dynamics*, ed. A. M. Anile and Y. Choquet - Bruhat (Berlin, Springer - Verlag).
- [53] Zakari, M. and Jou, D., 1993, Equations of State and Transport Equations in Viscous Cosmological Models, *Phys. Rev. D* **48**, 1597.
- [54] Romano, R. and Pavón, D., 1993, Causal Dissipative Bianchi Cosmology *Phys. Rev. D* **47**, 1396.
Pavón, D. and Zimdahl, W., 1993, Dark Matter and Dissipation, *Phys. Lett.* **179A**, 261.
- [55] Shandarin, S. F. and Zeldovich, Ya. B., 1989, The Large Scale Structure of the Universe: Turbulence, Intermittency, Structures in a Self Gravitating Medium, *Rev. Mod. Phys.* **61**, 185.
- [56] Szebehely, V. G., 1983, Gravitational examples of non deterministic dynamics, *Long-Time Prediction in Dynamics* ed. C W Horton, L E Reichl and V G Szebehely (New York: John Wiley), p. 227.
- [57] Calzetta, E., 1986, The Behavior of the Effective Gravitational Constants for Broken $SU(5)$, *Ann. Phys. (N.Y.)* **166**, 214.
- [58] Simon, J. Z., 1990, Higher Derivative Lagrangians, Nonlocality, Problems, and Solutions, *Phys. Rev. D* **41**, 3720
Parker, L. and Simon, J. Z., 1993, Einstein Equation with Quantum Corrections Reduced to Second Order, *Phys. Rev. D* **47**, 1339.
- [59] Calzetta, E. and Hu, B. L., 1989, Dissipation of Quantum Fields from Particle Creation, *Phys. Rev. D* **40**, 656.
- [60] Parker, L., 1968, Particle Creation in Expanding Universes *Phys. Rev. Lett.* **21**, 562
Parker, L., 1969, Quantized Fields and Particle Creation in Expanding Universes. I *Phys. Rev.* **183**, 1057
Parker, L., 1971, Quantized Fields and Particle Creation in Expanding Universes. II *Phys. Rev. D* **3**, 346.
- [61] Calzetta, E. and Sakellariadou, M., 1993, Semiclassical Effects and the Onset of Inflation, *Phys. Rev. D* **47**, 3184.
- [62] Hu, B. L., Paz, J. P. and Zhang, Y., 1993, Quantum Origin of Noise and Fluctuations in Cosmology, *Origin of Structure in the Universe*, ed. E. Gunzig and P. Nardone (Dordrecht, Kluwer) p 227.
- [63] E. Calzetta and B. L. Hu, 1993, Decoherence of Correlation Histories, in *Directions in General Relativity*, Vol. 2, ed. B. L. Hu and T. A. Jacobson (Cambridge, Cambridge University Press) p. 38.
E. Calzetta and B. L. Hu, 1993, Noise and Fluctuations in Semiclassical Gravity, *U. of Maryland preprint*.
- [64] Hartle, J. B., 1993, The Quantum Mechanics of Closed Systems, in *General Relativity and Gravitation 1992*, ed. R. J. Gleiser, C. N. Kozameh and O. M. Moreschi (Bristol, IOP)p. 81.
- [65] Starobinsky, A. A., 1986, Stochastic De Sitter (Inflationary) Stage in the Early Universe, in *Field Theory, Quantum gravity and Strings*, ed. N. Sánchez and H. de Vega (Heidelberg, Springer - Verlag).
- [66] Brandenberger, R., Feldman, H., Mukhanov, V. and Prokopec, T., 1993, Gauge Invariant Cosmological Perturbations: Theory and Applications, in *Origin of Structure in the Universe*, ed. E. Gunzig and P. Nardone (Dordrecht, Kluwer) p 13.

- [67] Cross, M. C. and Hohenberg, P. C., 1993, Pattern Formation outside of Equilibrium, *Rev. Mod. Phys.* **65**, 851.
- [68] Vilenkin, A., 1983, *Phys. Rev. D* **27**, 2848.
Halliwell, J. J., 1993, The Interpretation of Quantum Cosmological Models, in *General Relativity and Gravitation 1992*, ed. R. J. Gleiser, C. N. Kozameh and O. M. Moreschi (Bristol, IOP)p. 63.
- [69] Halliwell, J. J., 1987, *Phys. Rev. D* **36**, 3626.
- [70] Habib, S., 1990, Classical Limit in Quantum Cosmology: Quantum Mechanics and the Wigner Function, *Phys. Rev. D* **42**, 2566.
- [71] Paz, J. P. and Sinha, S., 1991, Decoherence and Back Reaction: the Origin of the Semiclassical Einstein Equations, *Phys. Rev. D* **44**, 1038.
Hu, B. L., Paz, J. P. and Sinha, S., 1993, Minisuperspace as a Quantum Open System, in *Directions in General Relativity*, Vol. 1, ed. B. L. Hu, M. P. Ryan and C. V. Viveshwara (Cambridge, Cambridge University Press) p. 145.
- [72] Berry, M., 1983, Semiclassical mechanics of regular and irregular motion, *Chaotic Behavior of Deterministic Systems* ed. G Iooss, R H G Helleman and R Stora (New York: North-Holland), p. 171.
- [73] Ashtekar, A., 1991 *Lectures on Non Perturbative Canonical Gravity*, (Singapore, World Scientific).
- [74] Ashtekar, A. and Pullin, J., 1990, Bianchi Cosmologies, a new Description, *Proc. Israel Phys. Soc.* **9**, 65.
- [75] Capovilla, R., Dell, J. and Jacobson, T., 1993, The Initial Value Problem in Light of Ashtekar's Variables, in *Directions in General Relativity*, Vol. 2, ed. B. L. Hu and T. A. Jacobson (Cambridge, Cambridge University Press) p. 66.
- [76] Temam, R., 1988, *Infinite Dimensional Dynamical Systems in Mechanics and Physics* (New York, Springer - Verlag).

MIXING PROPERTIES OF COMPACT $K = -1$ FLRW MODELS

George Ellis^{†‡} and Reza Tavakoli[†]

[†]Astronomy Unit, Queen Mary & Westfield College
Mile End Road, London. E1 4NS. UK

[‡]Department of Applied Mathematics, University of Cape Town
Rondebosch, C.P. 7700, South Africa
and SISSA, Miramare, Trieste, Italy

Abstract. We study the mixing properties of compact $k = -1$ FLRW models, as a function of the cosmological parameters and the topological compactification scale. We find the mixing to be less pronounced than some of the claims made previously, nevertheless in low density universes the mechanism can give rise to appreciable mixing and result in a reduction in the measured cmwbr anisotropy on a range of angular scales. These models also have other important features, namely: (i) they allow chaos to be expressed in a gauge invariant way; (ii) they are structurally stable and (iii) in low density universes they result in radically different estimates of length scales with potentially important consequences for the interpretation of the angular variation of the background radiation.

1. INTRODUCTION

Possible chaotic behaviour of the Einstein field equations has been invoked as a potential mixing mechanism responsible for some of the smooth features of the universe such as the observed isotropy of the microwave background radiation. Models so far studied in this connection, at least within the cosmological context, fall into a number of categories, among them: (I) anisotropic models, such as the Bianchi IX models [1, 2]; (II) compact Friedmann-Lemaître-Robertson-Walker (FLRW) models with negative spatial curvature [3, 4] and (III) FLRW models with positive curvature conformally coupled

to a massive scalar field [5]. There are, however, a number of shortcomings that most of these attempts share, namely very few of these models are (i) chaotic in the strict sense of the word; (ii) realistic, in the sense of being effective in the lifetime of the universe (iii) structurally stable, in the sense of remaining robust under small physically motivated perturbations and (iv) allow a gauge-invariant formulation of chaos.

Here we confine ourselves to (II) and start by recalling the well-known result from dynamical systems theory that for geodesic flows on constant curvature spaces to be chaotic (or more precisely Bernoulli) one requires two properties: firstly the curvature of the space needs to be negative and secondly the space needs to be compact. To get a feel for this, we first of all recall the geodesic deviation equation in the form

$$\frac{D^2}{ds^2}\zeta^\alpha = -R^\alpha_{\beta\gamma\delta}V^\beta\zeta^\gamma V^\delta \quad (1)$$

where Greek indices take values 0 to 3, ζ^α is the separation vector between neighbouring geodesics and V^β is a tangent to the geodesics. For the square of the normal component of ζ^α we have [9, 7, 6]

$$\frac{d^2\|\zeta_\perp\|^2}{ds^2} = -2K\|\zeta_\perp\|^2 \quad (2)$$

where $\|\zeta_\perp\|^2 = g_{\alpha\beta}\zeta_\perp^\alpha\zeta_\perp^\beta$ and K is the curvature of the 2-elements. This immediately shows that in all regions where $K < 0$, geodesics diverge exponentially. For spaces with constant negative curvature this is true for all 2-elements and therefore one has sensitive dependence on initial conditions (SDIC) [9, 8], i.e. neighbouring geodesics diverge exponentially. Note also that the rate of divergence in this case - the so called leading positive Lyapunov exponent - is proportional to $\sqrt{|K|}$ ¹ which, in view of the fact that K is a constant scalar for spaces with constant curvature, makes the characterization of the SDIC for such spaces gauge invariant. Secondly, in order to have mixing one also requires compactness. More precisely, Sinai [12] has shown that only subject to compactness are the geodesic flows on such negative curvature manifolds a k -flows - i.e. have positive Kolmogorov entropy K_e . He further shows that K_e is related to the volume V through $K_e \propto 1/V^{\frac{1}{N}}$, where N is the dimension of the manifold ($N = 3$ for spacetime). The requirement for compactness then follows immediately. It is perhaps instructive to recall here that this subject has a very old history dating back to Hadamard [10] (see also [9, 11, 8]).

Now for these ideas to be applicable to the case of geodesic flows in $k = -1$ FLRW models, it is necessary to show that a geodesic flow in the four dimensional spacetime induces a corresponding geodesic flow on the three dimensional negative curvature hypersurfaces of homogeneity. This was achieved by Lockhart et al [3], subject to a condition on the affine parameter. Briefly, the idea is to start with the FLRW in comoving coordinates with

$$g_{00} = -1, \quad g_{ij} = R^2(t)\gamma_{ij}, \quad (3)$$

¹We note here that the factor 2 present in equation (2) arises as a consequence of the fact that this equation represents the evolution equation for the square of the normal component of the deviation vector. We do not include this factor because we, in line with the usual practice in cosmology, deal with the vector itself rather than its square. To obtain results corresponding to the squared magnitude of the vector, the Lyapunov exponent adopted here would need to be scaled by a constant factor of $\sqrt{2}$.

where lower case Latin indices take values 1 to 3, R is the scale factor and γ_{ij} is the metric of the 3-space, together with the geodesic equation

$$\frac{d^2x^\alpha}{d\sigma^2} + \Gamma^\alpha_{\beta\gamma} \frac{dx^\beta}{d\sigma} \frac{dx^\gamma}{d\sigma} = 0, \quad (4)$$

and show that the spatial coordinates $x^i(\sigma)$ also satisfy geodesic equations with respect to the connections of the γ_{ij} . Rewriting this in terms of the cosmological time t and choosing an affine parameter $\lambda = \lambda(t)$ it can be shown that this can be done if the affine parameter λ is chosen such that it satisfies

$$\ddot{\lambda} + \frac{\dot{R}}{R} \left(2 - g_{ij} \frac{dx^j}{dt} \frac{dx^k}{dt} \dot{\lambda} \right) = 0. \quad (5)$$

In this way then there is a correspondence between the geodesic flow in the spacetime and the flow on the hypersurfaces of homogeneity.

Now if the hypersurfaces of homogeneity (i) possess negative constant curvature and (ii) are compactified without effecting the metric [13, 14, 15], then the above results show that geodesic flows on such spaces are strongly chaotic, with non-zero Kolmogorov entropy [12]. On this basis, Lockhart et al [3] argued, without in fact giving any concrete estimates, that this procedure gives “a natural mechanism leading to homogeneity and (local) isotropy in such a universe”. What is not made explicit here is that this is an “all time” statement and that in finite time, even if mixing is present, it is partial.

We should emphasize here that geodesic flows on compact spaces of constant negative curvature have important properties often absent in other chaotic examples so far found in cosmology. In particular, the SDIC (or more precisely the Lyapunov exponent) in this case is (i) expressible in terms of curvature scalar and is therefore gauge invariant (see [19] for a discussion of the problems that arise in the case of Bianchi IX), (ii) effective at all epochs, even though it is much more pronounced in the earlier epochs with high curvatures. This is in contrast to the Bianchi IX ‘mixing’ which is essentially only effective at very early times when quantum mechanical effects become important [20]; and (iii) structurally stable [9], which is important in view of the likelihood that cosmological models may be fragile [21, 22]. Consequently, if these models produce significant mixing in the lifetime of the universe, they would also have other nice features.

Recently, Gurzadyan and Kocharyan [4] have given an estimate of the stretching in such models. They claim that the effect is extremely large and could decrease the initial perturbations in the photon gas by factors ranging from 10^{-137} to 10^{-3} . Their analysis, however, suffers from a number of shortcomings, namely: (a) the assumption that the functional form for the scale factor $R(t)$ for the case of $k = -1$ is given by $R(t) \propto t^\alpha$, $\alpha = \text{const}$ ($\frac{1}{2} \leq \alpha \leq \frac{2}{3}$); (b) lack of an explicit dependence on the content of the universe such as the density parameter Ω_0 ; (c) absence of compactification scale in their calculations; and (d) unfounded assumptions implicitly made regarding the behaviour of geodesics and their relation to the observed radiation temperature anisotropies.

Now in view of the shortcomings in [3, 4], our aim here is to take a closer look at the instability properties of the geodesic flows in $k = -1$ FLRW cosmology and give a concrete estimate of their mixing effects in terms of observable parameters. This is mainly based on results reported in [25].

2. THE MAIN EFFECTS

The main effects of the geodesic instability in the negatively curved FLRW models can be understood in terms of three separate effects: (1) the *projection effect*, which is due to the fact that the geodesics in the 4-dimensional FLRW space-time are projected (along the fundamental world lines defined by standard comoving coordinates) into 3-hypersurfaces of homogeneity; (2) the *spatial geometrical effect*, which amounts to a stretching due to the SDIC of the connecting vectors between neighbouring geodesics in each 3-hypersurface; and (3) the *topological effect*, which is related to the mixing caused by the compactification assumed (similar to the mechanism of mixing in automorphisms of the torus - the so called Arnold Cat Map [24]).

Regarding the relation between (1) and (2), it is instructive to bear in mind what actually happens on a particular 3-hypersurface of homogeneity. The projections of the null geodesics always diverge exponentially and therefore result in stretching orthogonal to the geodesics. On the other hand the projection of the fundamental particles will correspond to stationary points. In this way one may clearly distinguish between timelike and null geodesics. Also since for the null geodesics all paths correspond to motion at the speed of light, there is no separation of points along the direction of flow. Furthermore, because of the spherical symmetry of the FLRW metrics about every point, mapping of images orthogonal to the flow is necessarily distortion-free. In the case of those timelike geodesics corresponding to realistic particle motions in space-time, the motion of the projected points in the 3-space is negligible.

Finally, the topological effect (3), can only occur when the diameter of \mathcal{I} (the intersection of the past light cone C_{-p} at the point of observation p with the surface of last scattering Σ , given by $t = t_d$) is larger than the compactification scale L_c of the 3-space. In that case, \mathcal{I} can cross Σ many times and have many self-intersections in Σ .

It is the combination of these self-intersections coupled with the SDIC which is, in the all-time limit, chaotic with a positive Kolmogorov entropy and which could approach ‘chaotic’ behaviour within the finite age of the universe if the length scales involved are appropriate. Clearly then any ‘mixing’ caused by topology is directly related to the compactification scale.

3. THE PROJECTION AND GEOMETRICAL EFFECTS

The stretching effects for the null geodesics, due to the projection and geometric effects may be calculated in two ways. We consider these briefly in this section.

3.1. Use of radial spatial distance

The first procedure employs the radial spatial distance d , which in a constant time hypersurface $t = \text{const}$ where $t_0 \leq t \leq t_d$ (the suffixes 0 and d denote the present time and the time of the decoupling respectively), is defined as:

$$d(z) = R(t)\lambda(z) = \frac{1}{(1+z)}\lambda(z)R_0, \quad (6)$$

where $R_0 = R(t_0)$, the redshift z is determined by $R_0/R(t) = (1+z)$, and λ is the coordinate distance given by

$$\lambda(z) = \int_{t_0}^t \frac{dt}{R(t)} = \int_0^z \frac{dz}{R'(z)(1+z)}. \quad (7)$$

To determine λ , we require \dot{R} which is obtained from the Friedmann equation

$$\frac{\dot{R}^2}{R^2} = \frac{\kappa\rho}{3} - \frac{k}{R^2}. \quad (8)$$

Now in presence of a non-interacting combination of matter and radiation, the total density takes the form

$$\rho = \frac{A}{R^3} + \frac{B}{R^4} \quad (9)$$

where the constants A and B are given by $A = \rho_{m0}R_0^{-3}$ and $B = \rho_{r0}R_0^{-4}$, with ρ_{r0} and ρ_{m0} representing the present values of the radiation and matter densities respectively. The Friedmann equation (8) evaluated at present gives

$$\frac{k}{R_0^{-2}} = H_0^{-2}(\Omega_0 - 1), \quad (10)$$

where the density parameter is

$$\Omega_0 \equiv \frac{\kappa\rho_0}{3H_0^{-2}} = \frac{\kappa(\rho_{m0} + \rho_{r0})}{3H_0^{-2}} \equiv \Omega_{m0} + \Omega_{r0},$$

and Ω_{m0} and Ω_{r0} are the present values of the matter and radiation density parameters respectively. Now confining ourselves to the cases where $k \neq 0$ we can employ the Friedmann equation (8) to substitute for \dot{R} in (7) which after integration yields:

$$\lambda(z) = \log \left[\frac{(2\sqrt{1-\Omega_0} + 2 - \Omega_0(1+e))(1+z)}{2\sqrt{(1-\Omega_0)(\Omega_0ez^2 + \Omega_0z(1+e) + 1) + \Omega_0z(1-e) - \Omega_0(1+e) + 2}} \right], \quad (11)$$

where

$$e = \frac{\rho_{r0}}{\rho_{m0} + \rho_{r0}} = \frac{\Omega_{r0}}{\Omega_{m0} + \Omega_{r0}} \approx \frac{\Omega_{r0}}{\Omega_{m0}}. \quad (12)$$

Now the stretching orthogonal to the geodesics in a surface $t = const$ considered by [4], as the basis of their conclusions, is given by

$$\delta = R(t) \exp(hd(z)), \quad (13)$$

where the constant h is the so-called Lyapunov exponent, which for the case of spaces of constant negative curvature is proportional to the square root of the curvature scalar, being in turn proportional to the inverse square of the scale factor. This results in $h = \frac{1}{R}$, which gives $hd(z) = \lambda(z)$ and therefore (13) becomes:

$$\delta = R(t) \exp(\lambda(z)). \quad (14)$$

This is a special solution of the geodesic deviation equation in the 3-space which does not correspond to the geodesics diverging from a point, as would be appropriate when considering the projection of the null geodesics generating our past light cone, which are the geodesics by which we receive the cosmic microwave background radiation. The appropriate solution, satisfying $z = 0 \Leftrightarrow \lambda = 0 \Leftrightarrow \delta = 0$ can be shown [25] to be of the form:

$$\delta(z) = \alpha \frac{R_0}{(1+z)} \sinh(\lambda(z)). \quad (15)$$

where the constant α is the angle between the geodesics in the 3-space at the point $\lambda = 0$ where they intersect, which is the same as the angle between the null geodesics as measured by the observer at the vertex of the null cone.

This expression enables us to quantify the effect of the stretching due to the geodesic instability in any surface $t = \text{const}$ (say $t = t_* \Leftrightarrow z = z_*$) by projecting into that surface along the (comoving) fundamental flow lines, using the projection factor $R(t_*)/R(t)$. For example, in the decoupling surface ($t = t_d \Leftrightarrow z = z_d$), we find that corresponding to the parameter distance $\lambda(z)$, the (projected) distance along the geodesics is $\bar{\delta}(z) = \delta(z)R_d/R(t) = \delta(z)(1+z)/(1+z_d)$ and the (projected) distance apart of the geodesics is

$$\bar{\delta}(z) = \delta(z)R_d/R(t) = \delta(z)(1+z)/(1+z_d) = \alpha \sinh(\lambda(z)) \frac{R_0}{(1+z_d)}, \quad (16)$$

with $\lambda(z)$ given by (11).

The maximum distance apart (which determines whether or not chaotic behaviour occurs) is given by evaluating this expression at $z = z_d$, giving the length scale

$$\bar{\delta}(z_d) = \alpha \frac{R_0}{(1+z_d)} \sinh(\lambda(z_d)). \quad (17)$$

An important feature of the expressions (15- 17) is that by using (10) with $k = -1$ together with (11) and (12), they can be expressed explicitly in terms of $\Omega_{0r}, \Omega_{0m}, H_0$ and z_d . Expression (17) gives the maximum spatial distance apart in the surface of decoupling that is reached by the projection of a pair of null geodesics making an angle α for an observer at t_0 .

Similarly, for projection into the present-day surface of constant time $t = t_0$ we obtain

$$D(z) \equiv \frac{R_0}{R(t)} \delta(z) = \alpha R_0 \sinh(\lambda(z)) \Rightarrow D(z_d) = \alpha R_0 \sinh(\lambda(z_d)). \quad (18)$$

where the final expression gives the present distance apart corresponding to the comoving maximum distance apart at the time of decoupling.

3.2. Use of observer area distance

It is instructive to compare the above estimate of the stretching with the usual estimate employed in cosmology. To this end recall the *observer area distance* r_0 defined through [26]

$$l(z) = r_0(z)\alpha \quad (19)$$

where α is the angle subtended by a (small) sphere of diameter l at redshift z (corresponding to time t and scale factor $R(t)$). Clearly the form of the expression for r_0 depends on the assumed content of the universe, as well as the value of Ω_0 , and in presence of both matter and radiation it can be expressed as [27]

$$r_0(z) = \frac{A}{(1+z)^2} \left\{ (q_0 - 1)[1 + 2q_0z + q_0z^2(1-\beta)]^{\frac{1}{2}} - (q_0 - q_0\beta z - 1) \right\} \quad (20)$$

where $q_0 \neq 0$ and the constant A is given by

$$A = \frac{1}{H_0 q_0 (q_0 + \beta - 1)}, \quad (21)$$

and the parameter β is related to the ratio of the present matter and radiation densities by

$$\frac{2\beta}{1-\beta} = \frac{\rho_{m0}}{\rho_{r0}} = \frac{\Omega_{m0}}{\Omega_{r0}} \Leftrightarrow \beta = \frac{\Omega_{m0}}{\Omega_{m0} + 2\Omega_{r0}}.$$

Thus a pure matter or radiation universe would correspond to $\beta = 1$ and $\beta = 0$ respectively.

The corresponding projected separation distance $\bar{\delta}(z)$ in the three space $t = t_d$ is then given by simple scaling with $R_d/R(t)$:

$$\bar{\delta}(z) = \frac{R_d}{R(t)} l(z) = \frac{1+z}{1+z_d} r_0(z)\alpha. \quad (22)$$

Note we use the same symbol here as in equation (17) since they both refer to the same quantity.

Recalling that $q_0 = \frac{1}{2}\Omega_0$, we obtain:

$$\bar{\delta}(z) = \frac{B}{1+z} \left\{ (\Omega_0 - 2)[1 + \Omega_0 z + \frac{\Omega_0}{2} z^2 (1 - \beta)]^{\frac{1}{2}} - (\Omega_0 - \Omega_0 \beta z - 2) \right\} \quad (23)$$

where

$$B \equiv \frac{2}{\Omega_0 H_0 (\Omega_0 + 2\beta - 2)(1+z_d)} \alpha. \quad (24)$$

This formula encodes all the dynamical effects including the exponential stretching due to the negative curvature, without explicit reference to concepts in nonlinear dynamical systems theory. An analogous expression for the distance $\bar{\delta}(z_d)$ (given by (17) of the previous section) may be obtained by evaluating (25) for $z = z_d$:

$$\bar{\delta}(z_d) = r_0(z_d)\alpha. \quad (25)$$

Similarly an alternative formula for $D(z)$ given by (18) may be obtained by projecting into the present-day surface of constant time $t = t_0$:

$$D(z) = \alpha r_0(z) \frac{R_0}{R(t)} \Rightarrow D(z_d) = \alpha r_0(z_d)(1+z_d). \quad (26)$$

Despite the apparent difference between the expressions (17 and 22) or (18 and 26), we shall see in section (5) that they are equivalent to one another in the sense that they lead to identical consequences.

4. TOPOLOGICAL AND ANGULAR EFFECTS

As was discussed in section (2), the presence and the degree of mixing depends crucially upon the nature of topology of the universe.

4.1. Topological effects

For significant mixing to take place as a result of topological compactification, the stretching in some surface $t = \text{const}$ must be much larger than the spatial compactification scale L_c in that surface, i.e.

$$\bar{\delta}(z_d) \gg L_c(t_d). \quad (27)$$

Alternatively, we may express this through the ‘re-entry factor’ f , defined in terms of area distance:

$$f = \frac{\bar{\delta}(z_d)}{L_c(t_d)} = \frac{r_0(z_d)}{L_c(t_d)}\alpha, \quad (28)$$

which gives a comparison between the compactification scale L_c and $\bar{\delta}(z_d)$ and which clearly depends on the angle α .

The question is whether there exist angles for which (27), which is equivalent to $f \gg 1$, can hold. To determine this we need to consider both the theoretical and observational constraints on L_c . Mathematically, there are two important topological features of the $k = -1$ three-spaces which are distinct from the $k = 0$ case. Firstly, it has been shown by Thurston [28] that the compact orientable hyperbolic $k = -1$ three-spaces are countably infinite in number (as opposed to $k = 0$ three-spaces which are finite). Secondly, as opposed to the $k = 0$ case, the compactification scale L_c for the $k = -1$ three-spaces cannot be chosen arbitrarily. It turns out that the volumes of all the complete compact hyperbolic three-spaces form a well ordered set with a lower bound (a manifold of minimum volume) [28]. This, however, is not a constructive result and therefore the manifold with the minimum volume is not known. So far the example with the smallest volume V has $V = 0.94R^3$ [29], and it has been shown (again non-constructively) that the lowest volume of $k = -1$ three-manifolds must be larger than $0.00082R^3$ [31].

Now recalling that a hyperbolic ball of radius $\mathcal{L} = R\chi$ has a volume of $V = \pi R^3(\sinh(2\chi) - 2\chi)$ [30], allows χ to be calculated to give $\mathcal{L} \approx 0.59R$. Assuming that $L_c \approx \mathcal{L}$, the compactification scale at present can be evaluated to be

$$L_c(t_0) \approx \mathcal{L} \approx \frac{1769}{\sqrt{1 - \Omega_0}} h^{-1} Mpc,$$

in the case of $V = 0.94R^3$ and

$$L_c(t_0) \approx \mathcal{L} \approx \frac{174}{\sqrt{1 - \Omega_0}} h^{-1} Mpc,$$

in the case of $V = 0.00082R^3$. In these expressions we have taken the current value of the scale factor to be $R_0 = \frac{2999}{\sqrt{1 - \Omega_0}} h^{-1} Mpc$ (with $0.4 < h < 1$).

Theoretically then the identification scale has to lie between these two limits.

Observationally, there are a number of estimated lower bounds for the the identification scale today. These range from orders of 200 to 500 Mpc (on the basis of galactic observations [33, 34]), to the order of $2400h^{-1} Mpc$ (based on the COBE anisotropies [35]). Other observations based on quasar images give an estimate of $600Mpc$ (see [35] for references). Observationally then we may safely assume that the minimum scale is greater than $200Mpc$.

The expressions for L_c and f derived above depend on the density parameter Ω_0 and the value of h , nevertheless, in the light of the minimum estimates for the L_c given above and from the numerical results in the next section, we shall see that (27) can be fulfilled for realistic values of Ω_0 and H_0 and mixing of the type envisaged in the section 2 can occur.

4.2. Angular effects

The angular formula (17,25) depend on Ω_0 and give noticeably different results in low and high density cases. Recalling that $r_0 = r_0(z, \Omega_0)$ (cf. 20), this difference may be quantified in terms of the ratio

$$F(z, \Omega_0) = \frac{r_0(z, \Omega_0)}{r_0(z, 1)}. \quad (29)$$

This shows the ratio between scale sizes in a universe with critical density parameter ($\Omega_0 = 1, q_0 = 1/2$) and a universe with density Ω_0 . Evaluated in the last scattering surface, this gives $\tilde{F} = F(z_d, \Omega_0)$, showing the ratio between the physical sizes sampled on the last scattering surface. We shall give an estimate of this quantity as a function of Ω_0 in the next section.

5. RESULTS

As an estimate of the (geometrical) stretching effects, we calculated the stretching $D(z)$ given in sections (3.1) and (3.2) projected onto the present time. Figures 1-3 show $D(z)$ as a function of z with $\Omega_{0r} = 2.56 \times 10^{-5} h^{-2}$ and $0.4 < h < 1$ where Ω_{0m} takes values 0.04, 0.4 and 1 respectively. As can be seen the stretching (i.e. the factor of $\sinh(\lambda(z))$) in equation (18) and its analogous implicit counterpart in equation (26)) increases dramatically as Ω_{m0} is decreased, but even for the maximal value of $\Omega_{m0} = 1$ the stretching is a factor of 2 to 5 (depending upon the value of h). For the minimal value of $\Omega_0 = 0.04$ the corresponding factors range from ≈ 34 to 73. Comparing the plots of the function $D(z)$ from (18) and (26) shows that despite their different forms they lead to identical results.

As an estimate of the topological effects, Figure 4 shows the plot of f (defined in section 4.1 above) as a function of Ω_0 with α chosen to be 10 degrees to make it compatible with the COBE angular diameter measurement and the present day identification scale taken to be $L_c(t_0) = 500$ Mpc (and $L_c(t_d) = 0.5$ Mpc). As can be seen $f \gg 1$, even for values of α smaller than ten degrees.

As a measure of the angular effects, we show in Figures 5 and 6 the plots of the $F(z, \Omega_0)$ (given in section 4.2) as a function of z for $\Omega_0 = 0.04$ and 0.4 respectively. As can be seen this effect decreases dramatically as Ω_0 is increased. The importance of this factor is that it indicates that the results of Stevens et al [35] (which may be taken as the best present estimates of the effect of compactness on cmwbr anisotropy) do not necessarily preclude an interesting ‘small universe’ in the low density case; for as a first estimate of the possible size allowed by the microwave background radiation anisotropy measures, we can take the scale estimated by Stevens et al of 2400 Mpc today, and use the factor \tilde{F} to estimate that (projected to present day length scales) this is compatible with an identification scale in a low density universe of $2400/\tilde{F}$. This can be as small as 140 Mpc, i.e. less than the estimate given on the basis of other observational criteria. For a discussion of the consequences of such mixing for the observed redshifts and the Planckian spectrum the reader is referred to [25].

Our results show that since decoupling, there has been enough time for exponential expansion of the projected null geodesics plus the topological effect to produce a significant mixing (albeit not as large as some of the estimates given previously [3, 4])

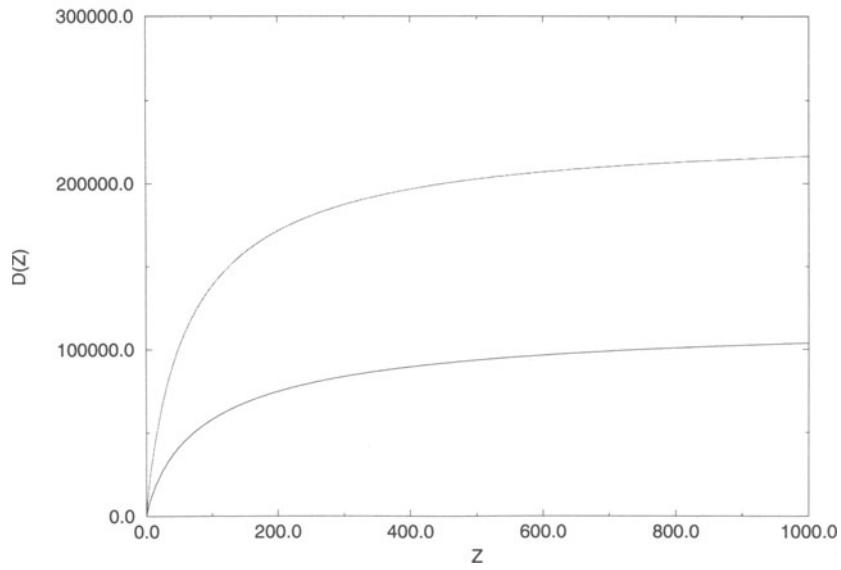


Figure 1. Plot of function $D(z)$ vs z with $\Omega_0 = 0.04$ and $h = 0.4, 1$.

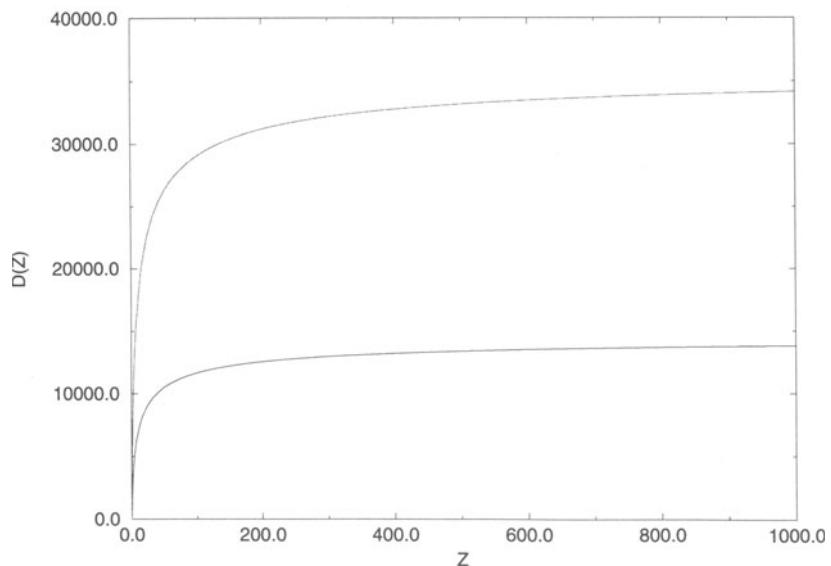


Figure 2. Same as Fig 1, but with $\Omega_0 = 0.4$

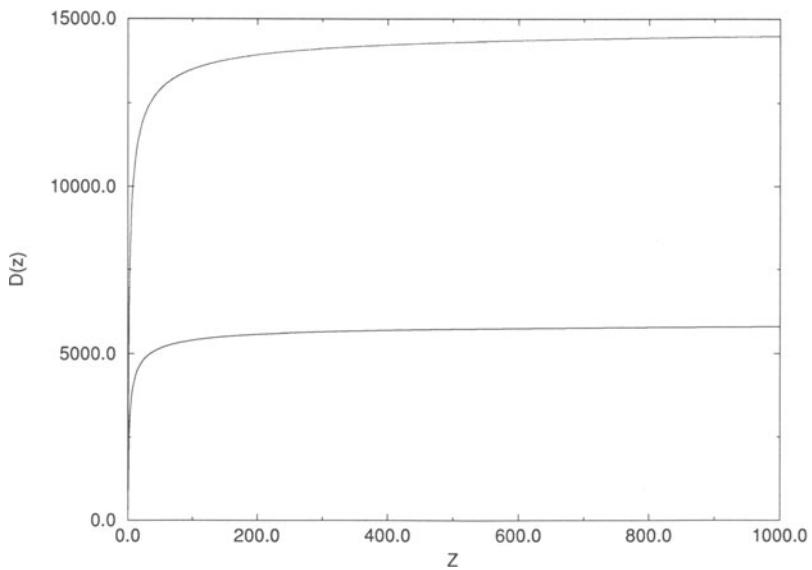


Figure 3. Same as Fig 1, but with $\Omega_0 = 1$.

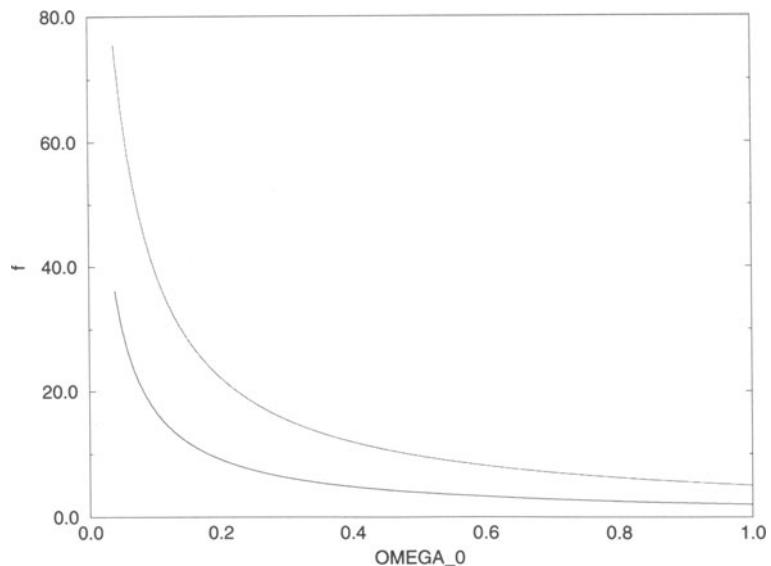


Figure 4. Plot of f vs Ω_0 with $\alpha = 10$ degrees and $L_c(t_d) = 0.5$ Mpc.

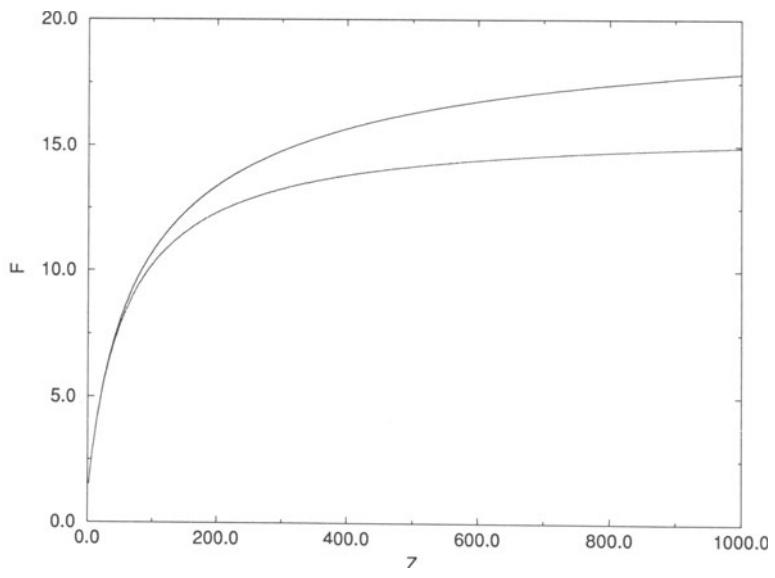


Figure 5. Plot of function $F(z, \Omega_0)$ vs z for $\Omega_0 = 0.04$ and $h = 0.4, 1$.

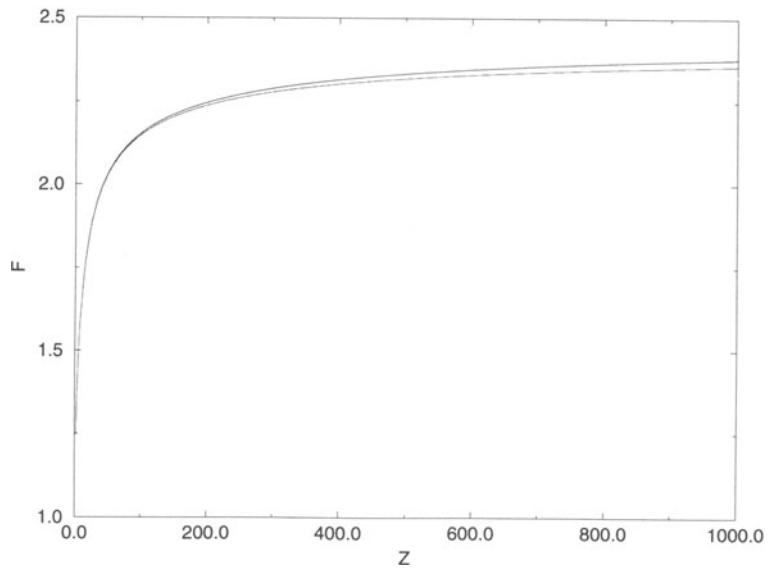


Figure 6. As in Fig 5, but with $\Omega_0 = 0.4$.

and so causing a reduction in cmbr anisotropy on some angular scales (depending on the identification length scales, and the parameters describing the universe model). Also this stretching can make a non-negligible difference to the length scales on the surface of last scattering in low density universes, which could alter estimates of cmbr anisotropy measurements predicted on various angular scales in such models.

Acknowledgments GE was supported by FRD (South Africa) and MURST (Italy) and RT by SERC UK Grant number H09454.

REFERENCES

- [1] See for example Belinski, V.A., Khalatnikov, I.M. and Lifshitz, E.M. (1970) *Adv. Phys.*, **19**, 525.; Barrow, J.D. (1982) *Phys. Rep.*, **85**, 1.; Burd, A.B., Buric, N. and Tavakol, R.K. (1991) *Class. Quantum Grav.*, **8**, 123.
- [2] Koiller, J., De Mello Neto, J.R.T. and Damiao Soares, I. (1985) *Phys. Lett.*, **110A**, 260.
- [3] Lockhart, C.M., Misra B. and Prigogine I. (1982) *Phys. Rev.*, **D 25**, 921.
- [4] Gurzadyan V. G. and Kocharyan A.A., (1992) *Astron. Astrophys.*, **260**, 16.14
- [5] Calzetta, E. and El Hasi, C. (1992): ‘Chaotic Friedmann Robertson Cosmology’, Preprint.
- [6] Zydowski, M. and Lapeta, A. (1990) *Phys. Lett.*, **148A**, 239.
- [7] Arnold, V.I. (1989) *Mathematical Methods of Classical Mechanics*, 2nd ed. (Springer-Verlag, New York).
- [8] Hopf, H. (1936) *Trans. Am. Math. Soc.*, **45**, 241.
- [9] Anosov, D. (1967) *Proc. Steklov Ins.*, **90**, Ins. No. 90.
- [10] Hadamard J. (1898) *J. Math. Pures et Appl.*, **4**, 27.
- [11] Hedlund, G.A. (1939) *Bull. Amer. Math. Soc.*, **45**, No. 4, 241.
- [12] Sinai, Y.G. (1960) *Sov. Math. Dok.*, **1**, 335.
- [13] Wolf, J.A. (1967) *Spaces of constant curvature*, McGraw-Hill, New York.
- [14] Thurston, W.P. and Weeks, J.R. (1984) *Scientific American*, July, p.94.
- [15] Ellis, G.F.R. (1971) *Gen. Rel. Grav.*, **11**, 11.
- [16] Chitre, D.M. (1972) University of Maryland Technical Report No. 72-125.
- [17] Pullin, J. (1990) Syracuse preprint 90-0732.
- [18] Ellis, G.F.R. and Rothman, T. (1993) ‘Lost Horizons’. *Am. Journ. Phys.* (to appear).
- [19] Burd, A. and Tavakol, R. (1993) *Phys. Rev.*, **D 47**, 5336.
- [20] Nielson, H.B. and Rugh, S.E. ‘Chaos in Fundamental Forces’, In *Proceedings of a symposium on Quantum Physics and the Universe*, Waseda University, 1992. To appear.
- [21] Tavakol, R.K. and Ellis, G.F.R. (1988) *Phys. Lett.*, **130A**, 217.
- [22] Coley, A.A. and Tavakol, R.K. (1992) *Gen. Rel. Grav.*, **25**, 835.
- [23] Schneider, P., Ehlers, J., and Falco, (1992). *Gravitational Lensing*. Springer, Berlin.

- [24] Arnold, V. and Avez, A. (1968) *Ergodic Problems of Classical Mechanics*, Benjamin, New York.
- [25] Ellis, G. and Tavakol, R. (1993) 'Geodesic Instability and Isotropy of CMWBR', *Class. Quantum Grav.*, to appear.
- [26] Ellis, G.F.R. (1971). In *General Relativity and Cosmology*, Proc Int School of Physics "Enrico Fermi" (Varenna), Course XLVII. Ed. R K Sachs (Academic Press), 104-179.
- [27] Matravers, D.R. and Aziz, A.M. (1988) *MAASSA*, **47**, 124.
- [28] Thurston, W.P. (1978) *The geometry and topology of three manifolds*, Princeton University Lecture Notes.
- [29] Matveev,S.V. and Fomenko, A.T. (1988) *Russ. Math. Surv.*, **43**,3; Weeks, J. (1985) Princeton University Ph.D. Thesis.
- [30] Gott J.R. III (1980) *Mon. Not. R. Astron. Soc.*, **193**, 153.
- [31] Meyerhoff, R. (1986) *Commun. Math. Helv.* **61**, 271.
- [32] Hayward, G. and Twamley, J. (1990) *Phys. Lett.*, **149 A**, 84.
- [33] Ellis, G.F.R. and Schreiber, G. (1986) *Phys. Letts.*, **115A**, 97.
- [34] Fairall, A.P. (1985) *Mon. Not. Ast. Soc. S. A.* **44**, 114.
- [35] Stevens, D, Scott, S. and Silk, J, (1993). 'Microwave Background Anisotropy in a Toroidal Universe'. To appear, *Phys Rev Lett*.

CLASSICAL AND QUANTUM CHAOS IN ROBERTSON-WALKER COSMOLOGIES

Roman Tomaschitz

Physics Department, University of the Witwatersrand,
Johannesburg, WITS 2050, South Africa

Abstract. An elementary review of my work on the physical impact of the topological structure of space-time is given. An account on classical chaos in an open, multiply connected universe is presented. The uniformity of the galactic background is related to the erratic behavior of the classical world lines around the chaotic nucleus of the universe. On the quantum level we discuss particle creation, backscattering, anisotropy in the microwave background, parity violation and how all this relates to the multiple connectivity of the open spacelike slices.

1. INTRODUCTION

If one agrees to do cosmology on the basis of a space-time continuum, a Riemannian four-manifold, one is soon confronted with the choice of the topology. What is the global topological structure of space-time? This question was immediately raised by the mathematician Felix Klein, when Einstein proposed his first cosmological model, but it was not until much later, that the topological impact on geodesic motion gained serious consideration [1, 2].

Is the universe finite or infinite? Most laymen, and I think also philosophers, would attach to the word “Universe” the attribute “infinite”. Many physicists however believe that finite models of the universe are handier for heuristic reasoning. In fact, the idea that the physical universe is infinite was rejected by Jordan on the grounds that we will never be able to look at infinity and verify what is happening there [3]. What he overlooks is that the global structure can manifest itself in (microscopic) physical phenomena.

We will see, cf. Sec. 3, that an infinite and multiply connected universe has a chaotic center [4, 5], that may well account for the equidistribution of the galaxies. Some kind of chaos is in my opinion necessary to achieve this, the actual problem here is not only to explain the uniformity, but rather the apparent deviations from perfect homogeneity.

The crucial question is not so much what *is* the topological structure of the universe, but rather *how does it evolve*. An infinite universe has indeed the ability to evolve [6], contrary to closed ones, whose time evolution is confined to a trivial rescaling of the length unit on the spacelike slices. In Sec. 5 we will discuss this evolution in terms of global metrical deformations of the spacelike slices, global in the sense that the local curvature stays constant. We show how simple coordinate transformations in the covering space can generate such deformations, and how they distort the center of the 3-slices. In Sec. 6 we show how this evolution provides a dynamical mechanism for particle creation [6, 7]. In Sec. 7 we demonstrate how such deformations generate an angular dependence in the temperature of the microwave background [7, 8]. In Sec. 8 we sketch how the topology of the universe leads to the non-conservation of parity [9], and to a possible explanation of the baryon asymmetry. Finally I mention here the fascinating possibility of a multiple connectivity of space-time in the small [10], and of particles emerging as topological excitations.

In Secs. 2 and 3 we give an introductory and more or less self-contained account on the geodesic problem in open hyperbolic 3-manifolds [11-14]. In Sec. 4 we discuss how this relates to chaos in RW-cosmologies[4, 5]. The quantum mechanical ground state problem for chaotic wave fields localized on the center of the 3-slices [4, 15], the dispersion phenomenon in RW-cosmologies, and horospherical flows [7,16] are reviewed in [17].

2. SOME ELEMENTARY GEOMETRICAL AND TOPOLOGICAL CONCEPTS

In this Section we sketch some methods to study the global behavior of geodesics on a 3-manifold in a quantitative way. With geodesic we mean here the shortest path between two points compared to all neighboring paths. The emphasis lies here on “neighboring”, because in a multiply connected space the geodesic variational problem has several local minima, if the two points lie sufficiently far apart. With geodesics we mean such local minima.

Global variational problems are in practice much harder to solve than local ones. In the case of multiply connected manifolds, the universal covering space construction provides a very efficient way to do that. The geodesics as we discuss here turn out to be the spacelike projections of the world lines in the RW-cosmologies discussed in Sec. 4. Let us start with an example.

The simplest example of a multiply connected hyperbolic 3-manifold is a solid torus, topologically the product of a finite interval and an annulus. It is best modeled in the Poincaré half-space H^3 , its universal covering space: the complex plane with a t -axis perpendicular to it, $t > 0$ (t denotes always a space coordinate), and a line element $d\sigma^2 = t^{-2}(|dz|^2 + dt^2)$. In this way we get an isometric copy of the Minkowski hyperboloid, i.e. hyperbolic space H^3 . The geodesics are either straight lines perpendicular to the complex plane or semicircles orthogonal to it. The totally geodesic planes are therefore either hemispheres on \mathbb{C} or Euclidean half-planes perpendicular to \mathbb{C} [12].

To construct the torus we place two concentric hemispheres with radii $r = 1$ and $r = |\alpha| > 1$, $\alpha \in \mathbb{C}$, onto the complex plane, see Fig. 1a, and identify them by the transformation $T_\alpha(z, t) = (\alpha z, |\alpha|t)$. This T_α leaves the metric on H^3 invariant. If α is positive the identification happens radially, if it is complex we rotate the hemispheres against each other before we identify them. The hyperbolic polyhedron F bounded by the two hemispheres and the annulus between their base circles in the complex plane is topologically a solid torus if we perform the above identification with T_α , see Figs. 1a,b. The metric $d\sigma^2$ of H^3 gets induced onto F , and (F, T_α) is thus a hyperbolic manifold.

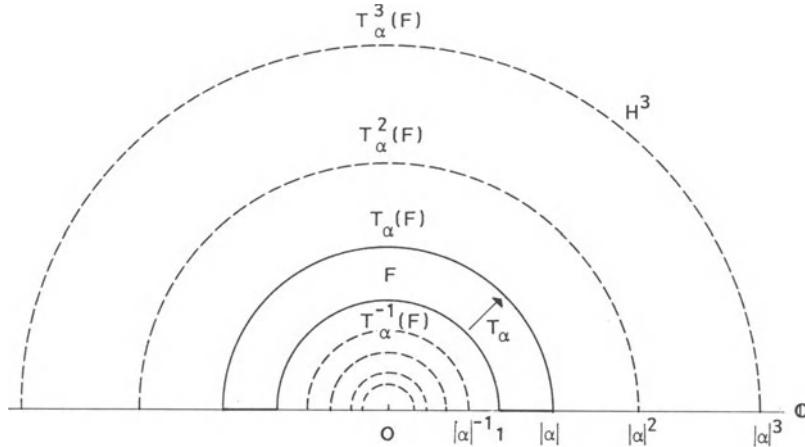


Figure 1a. Section through the Poincaré half-space H^3 . We place two concentric hemispheres of radii 1 and $|\alpha|$ onto the complex plane. F is the polyhedron bounded by these two hemispheres and the annulus between the base circles in \mathbb{C} . If we identify the two hemispheres radially (indicated by the map T_α), then (F, Γ) is topologically a solid torus, cf. Fig. 1b. The covering group Γ consists of all integer powers T_α^n . The images $T_\alpha^n(F)$ provide a tiling of H^3 , as well as of \mathbb{C} , the boundary at infinity of H^3 , with concentric hemispherical shells and annuli respectively. The tiles have two accumulation points in \mathbb{C} , namely 0 and ∞ . $\Lambda(\Gamma)$ is the limit set of Γ .

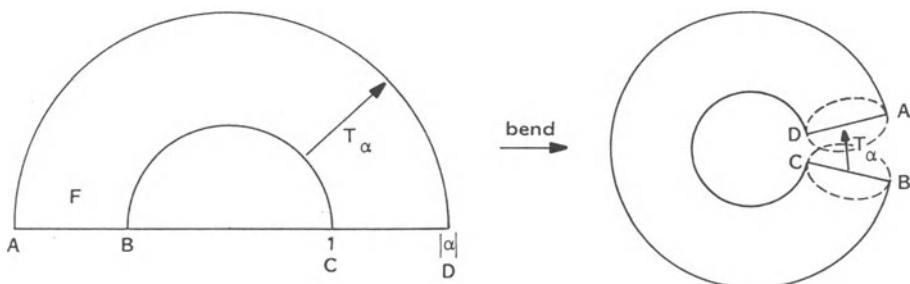


Figure 1b. If the polyhedron F is bent in the indicated way, it is quite obvious that (F, Γ) gives topologically a solid torus. Its boundary is the 2-torus obtained by identifying the boundary circles of the annulus in the complex plane. The hyperbolic metric gets singular at \mathbb{C} . Therefore (F, Γ) is an open 3-manifold.

One can now pose the question how many hyperbolic (i.e. of curvature -1) 3-manifolds with the topology of an open solid torus (the product of a finite open interval with an annulus) do exist. Sufficiently small coordinate patches on two hyperbolic manifolds can always be mapped isometrically onto each other, because they have the same curvature. Here however we must ask if there exists a global isometry, a diffeomorphism of a solid torus onto another, that respects the metric. The answer is that different α (in T_α) correspond to different, globally non-isometric (non-isotopic) solid tori.

In order that (F, T_α) is a Riemannian manifold, the induced metric has to fit smoothly on the identified hemispheres. There is a simple criterion for that. We define Γ , the covering group, as the set of all integer powers of T_α . Then the images $T_\alpha^n(F)$ of the polyhedron F under Γ must give a tiling of the covering space H^3 . This tiling is depicted in Fig. 1a. The tiles have two accumulation points, 0 and ∞ , the limit set $\Lambda(\Gamma)$. These accumulation points at infinity play an important part in analyzing the global behavior of geodesics, as well as in the spectral analysis of wave equations on the manifold [4, 8, 9].

By means of the covering group Γ and the polyhedral tiling $\Gamma(F)$ of H^3 we can project geodesics of the covering space H^3 into the 3-manifold (F, Γ) . Consider a geodesic s (semicircle orthogonal to the complex plane) in H^3 . It intersects a finite or infinite number N of polyhedral images, say $T_\alpha^{n_1}(F), \dots, T_\alpha^{n_N}(F)$. Evidently we can label them as adjacent. We denote the arc of s lying in $T_\alpha^{n_i}(F)$ by s_n . So we can regard s as the ordered sequence of arcs $\{s_{n_i}\}_{i=1,\dots,N}$. The projection of s into (F, Γ) we define as the ordered sequence $\{T_\alpha^{-n_i}(s_{n_i})\}_{i=1,\dots,N}$ of arcs in F , cf. Sec. 8. Initial and end points of adjacent arcs, lying on the concentric hemispheres bounding F , are identified by T_α , so that we obtain a smooth geodesic in the 3-manifold (F, Γ) . Every geodesic can be realized in this way. The whole works because the covering group Γ leaves the metric of H^3 invariant. We get so a perfectly quantitative realization of geodesic motion in the 3-manifold. The qualitative behavior of a geodesic depends very much on whether the initial and end points of the covering geodesic s are in the limit set $\Lambda(\Gamma)$, see Figs. 2-4.

3. THE CHAOTIC CENTER OF AN INFINITE, MULTIPLY CONNECTED 3-MANIFOLD

There are two generic classes of open hyperbolic 3-manifolds, namely solid handlebodies of genus $g \geq 1$ (the product of a finite open interval and a disc with some smaller discs removed), and thickened Riemann surfaces of genus $g \geq 2$, (the product of a finite open interval and a sphere with g handles attached). All that what has been said about the solid torus in Sec. 2 carries easily over to these cases [13, 14], and so we sketch them very shortly here.

The covering space constructions are indicated in Figs. 5-7, to which we refer in the following. The 3-manifolds (F, Γ) are again represented by a polyhedron F whose faces are identified in pairs by transformations T_i that leave the H^3 -metric invariant. The covering group Γ is now the discrete group consisting of all words with letters $T_i^{\pm 1}$. If we apply Γ to F we get a tiling of H^3 , which means that the Γ -images of F (apart from the identity) fill the interior of the hemispheres that bound F , cf. Figs. 5a and 6a. These images have accumulation points in the complex plane, like 0 and ∞ in Fig. 1a. In the case of Fig. 5 they constitute a Cantor set $\Lambda(\Gamma)$, totally disconnected and dense in itself, but not self-similar [18], in Fig. 6a they are a closed, fractal curve. The

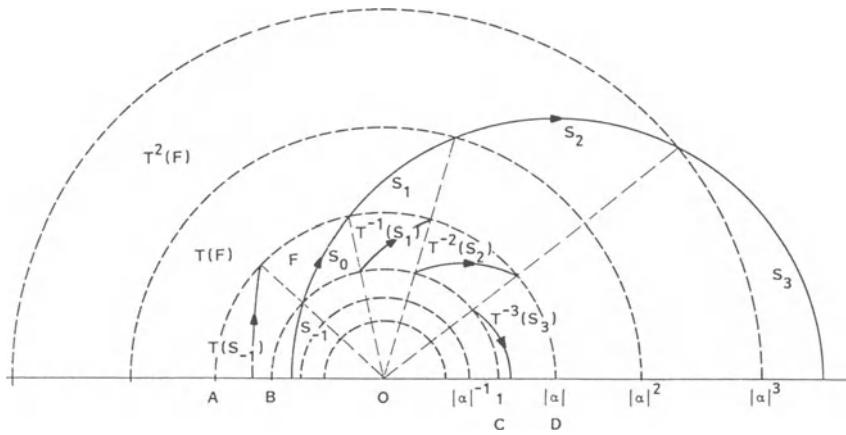


Figure 2a. The covering space construction, namely the representation of a 3-manifold as a polyhedron F with face-identification, is the appropriate means to study quantitatively geodesic motion on a multiply connected manifold. Every trajectory in (F, Γ) can be obtained by projecting a covering trajectory (semicircle in H^3) into the polyhedron F . The arc s_n , lying in the tile $T_\alpha^n(F)$, is mapped into F by $T_\alpha^{-n}(F)$. The trajectory in the 3-manifold appears now as the ordered sequence of arcs $(T_\alpha(s_{-1}), s_0, T_\alpha^{-1}(s_1), T_\alpha^{-2}(s_2), T_\alpha^{-3}(s_3))$. The initial and end points of the arcs are identified by T_α , as indicated by the hatched rays.

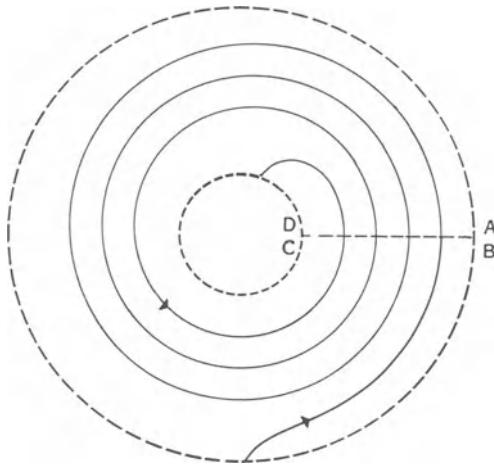


Figure 2b. The topology of the trajectory in Fig. 2a.

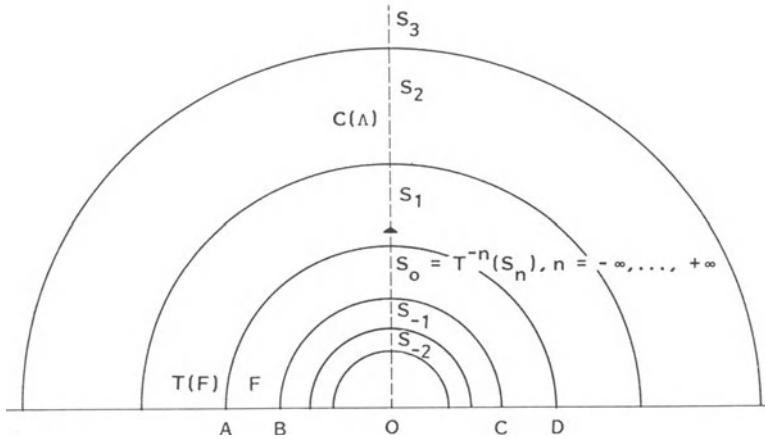


Figure 3. The covering trajectory is here a straight line $(s_{-\infty}, \dots, s_0, \dots, s_\infty)$, connecting the two limit points 0 and ∞ . Its projection into F is s_0 , infinitely covered by the images $T_\alpha^{-n}(s_n)$. s_0 is the only closed geodesic loop. It introduces a length scale in this infinite space. The hyperbolic length of this loop is $\log |\alpha|$.

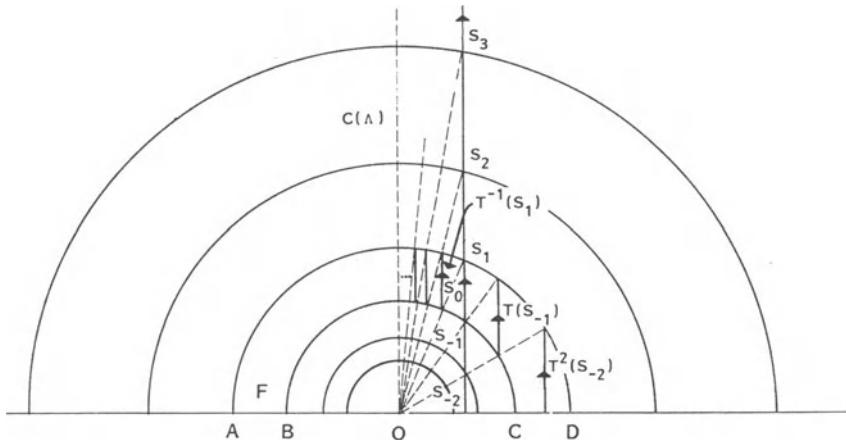


Figure 4a. The covering trajectory $(s_{-2}, \dots, s_0, \dots, s_\infty)$ has only one end point, ∞ , in the limit set. The projected arc pieces $T_\alpha^{-n}(s_n)$ accumulate at the limit cycle, cf. Fig. 3.

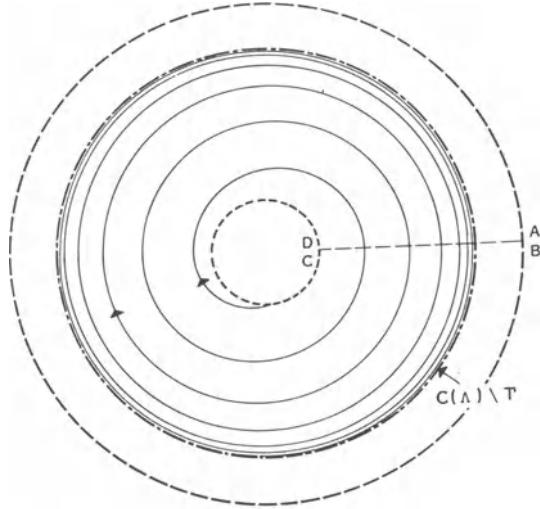


Figure 4b. The topology of the trajectory in Fig. 4a. Starting at infinity, it loops asymptotically into the limit cycle.

tiling $\Gamma(F)$ of H^3 induces a tiling of the complex plane, which can be used to determine the limit set $\Lambda(\Gamma)$, cf. Figs. 6b and 7a-c.

Geodesic motion on these multiply connected manifolds (F, Γ) can be realized in the same way as in the case of a solid torus, cf. Figs. 1-4, by projecting a H^3 -geodesic with the covering group Γ into F .

Let us look on the qualitative behavior of trajectories, keeping in mind Figs. 1-4. What is the analogue to the closed geodesic loop in Fig. 3? Let us start with the 3-manifold that has as limit set $\Lambda(\Gamma)$ a circle, cf. Fig. 6b. Consider the set of all the covering geodesics that have initial and end points on this circle. These arcs cover the hemisphere $C(\Lambda)$ placed on the circle. This hemisphere intersects all the polyhedral faces erected on the base circles in Fig. 6a. Consider the region of this hemisphere that lies above the polyhedral faces. We identify its boundary arcs, which lie on the polyhedral faces, as indicated in Fig. 6a. In this way we obtain a closed surface of genus two, a doughnut. This Riemann surface embedded in the 3-manifold is the analogue to the closed loop of Fig. 3. Clearly this surface has finite area, a metric of constant curvature -1 is induced on it from H^3 : it is the center $C(\Lambda) \setminus \Gamma$ of the open 3-manifold (F, Γ) .

Bounded geodesics in (F, Γ) have covering geodesics which have initial and end points in the limit set (as is the case in Fig. 3). It is easy to see that all these bounded geodesics lie in the center $C(\Lambda) \setminus \Gamma$. Geodesics on such a compact surface have all kinds of chaotic properties, almost all are densely filling the surface, some are closed loops.

We ask what happens with a trajectory that has a covering trajectory whose initial and end points are not in $\Lambda(\Gamma)$, but very close to it. The corresponding geodesic in the 3-manifold will still tend from infinity to infinity, as in Fig. 2, but it will loop a long time in a region close to the center. Such geodesics are not dense or ergodic or mixing or whatever in the strict mathematical sense, but in any practical way they will appear as such, provided that the end points of the covering geodesic are close enough to the limit set. This is important, for in my opinion some kind of erratic

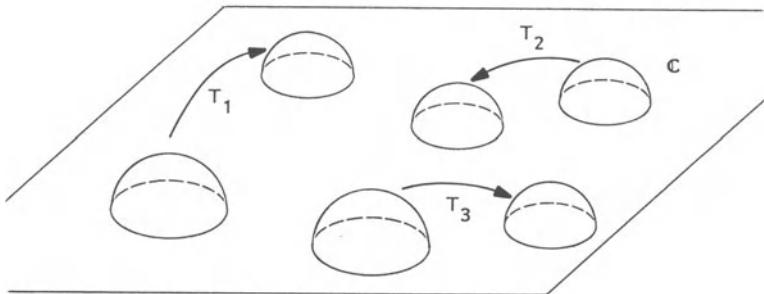


Figure 5a. Covering space construction of a handlebody. We choose $2g$ mutually disjointed hemispheres and identify them in pairs with elements T_i of the invariance group of the hyperbolic metric, so that T_i maps the outside of one sphere onto the interior of the other. The polyhedron F is now the space above the hemispheres and the complex plane. The polyhedral faces are the hemispheres, and there is one face at infinity of H^3 , analogous to the annulus in Fig. 1, namely the compactified complex plane with the $2g$ discs under the hemispheres removed.

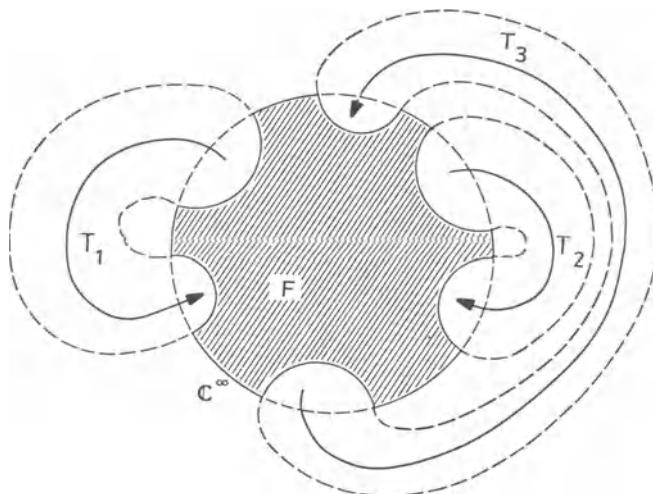


Figure 5b. The topology of the 3-manifold (F, Γ) in Fig. 5a. We compactify H^3 to a ball. Its boundary at infinity is the Riemann sphere. The polyhedron F is the hatched region. We identify the spherical caps as indicated. Topologically we attach in this way three solid handles to the ball.

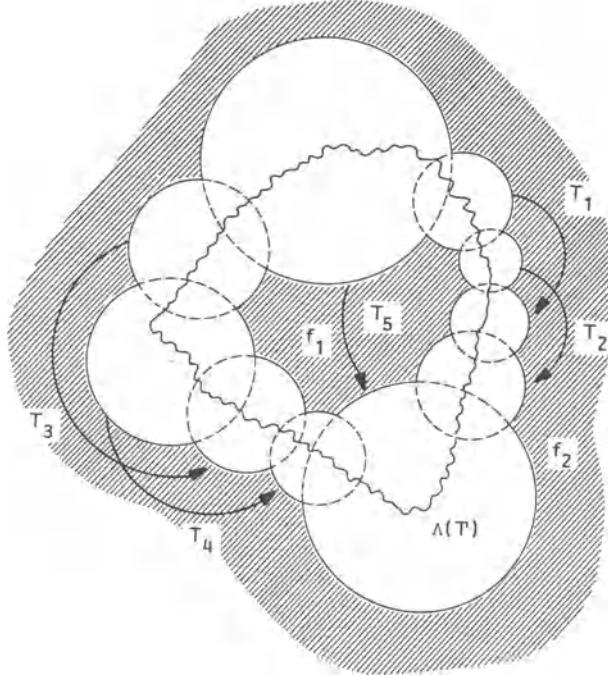


Figure 6a. Covering space construction for a thickened Riemann surface of genus two. Indicated are the base- circles of the hemispheres in the complex plane. Three-manifolds of this type can be realized by a ring of intersecting hemispheres with a suitable identification (T_i). The polyhedron F comprises the space above the hemispheres and the two faces f_1 and f_2 (hatched) at infinity. There is an analogue to Fig. 5b.

geodesic behaviour is necessary to account for the more or less uniform distribution of the galaxies. This uniformity does not seem to be perfect at all, and what is provided here is just a mechanism of imperfect classical chaos to achieve that.

The example in Fig. 6b is not yet generic for an infinite hyperbolic 3-manifold. Such a manifold has a limit set of non-integer Hausdorff dimension, cf. Figs. 7a-c. In this case the center of the 3-manifold is a three-dimensional finite domain: we start with the collection of all semicircles orthogonal to the complex plane which have their initial and end points in the fractal curve, and construct the hyperbolic convex hull $C(\Lambda)$ of them. This is now a three-dimensional domain. We consider the part of this domain that lies above the hemispheres in Fig. 6a, and identify its boundaries as indicated. We obtain a three-dimensional finite domain in the 3-manifold, which is itself not a manifold because it is pleated [13] due to the fractal nature of the limit set. The scenario concerning geodesic motion is otherwise quite similar to the foregoing. The bounded chaotic trajectories lie in this center, and nearly chaotic trajectories loop close to it.

4. GEODESIC MOTION IN MULTIPLY CONNECTED RW-COSMOLOGIES

These cosmologies are topologically the product of a hyperbolic 3-manifold and a time axis, $R^{(+)} \times (F, \Gamma)$, the time axis $R^{(+)}$ may be infinite, semi-infinite or whatever. The line element in the covering space $R^{(+)} \times H^3$ is $ds^2 = -c^2 d\tau^2 + a^2(\tau) d\sigma^2$, with $d\sigma$ as in Sec. 2. The corresponding metric we denote by $g_{\mu\nu}^{RW}$, the a^2 -scaled Poincaré metric on

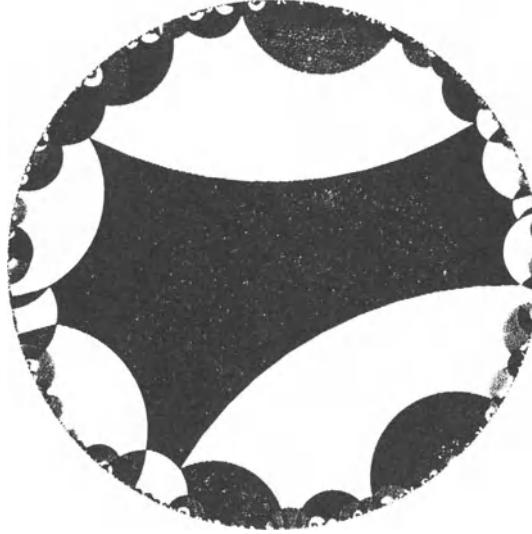


Figure 6b. A tiling on the boundary of H^3 that is induced by the tiling $\Gamma(F)$. The domain in the middle is the hatched domain f_1 in Fig. 6a, bounded by circular arcs on the base circles. The accumulation points of the tiles define in this case a circle $\Lambda(\Gamma)$, corresponding to the closed curve indicated in Fig. 6a. All tiles are Γ -images of f_1 .

the spacelike sections we denote in the following by g_{ij}^P . The metric $g_{\mu\nu}^{RW}$ gets induced on $R^{(+)} \times (F, \Gamma)$. Solving the geodesic equations in $R^{(+)} \times H^3$, we realize that the spacelike projections of the world lines are the covering geodesics described in Secs. 2 and 3. Their projections into (F, Γ) inherit the time parametrization of the covering trajectories. A covering trajectory will not reach the boundary at infinity of H^3 within a finite time. That means that its projection into the center of the 3-slices can be dense at best in the limits $\tau \rightarrow 0, (-\infty), +\infty$, i. e. backwards in time or at the end of the expansion. It may also happen, depending on the expansion factor a , that the covering trajectory does not reach the boundary even in these limits. That means that the actual covering trajectory is only a finite arc, well separated from the complex plane. If we project this arc into the center $C(\Lambda) \setminus \Gamma$, we get a trajectory of finite length. This trajectory may come close to every point in the center, but it is not dense in the strict sense.

Summing up, there is a large proportion of trajectories, namely those whose covering geodesics have end points close to the limit set (this set can have a dimension close to two), which will spend a long time in a domain close to the center of the 3-manifold, looping around there in an erratic way. This can be a mechanism to generate the current imperfect uniformity of the galactic background [6, 8].

Concerning Einstein's equations in this context [19], we mention that the curvature tensor of the covering space projects as it stands onto $R^{(+)} \times (F, \Gamma)$, likewise the energy-momentum tensor. This is so because their dependence on the space coordinates is only via the Poincaré metric, and thus both tensors are invariant with respect to the covering group. There is always the same well known relation between pressure, density, and the expansion factor, independent of the topology. Therefore it seems to me difficult to gain information about the evolution of the universe on a global level by means of Einstein's equations. In my view it is also a mistake to rely on the positivity arguments for pressure

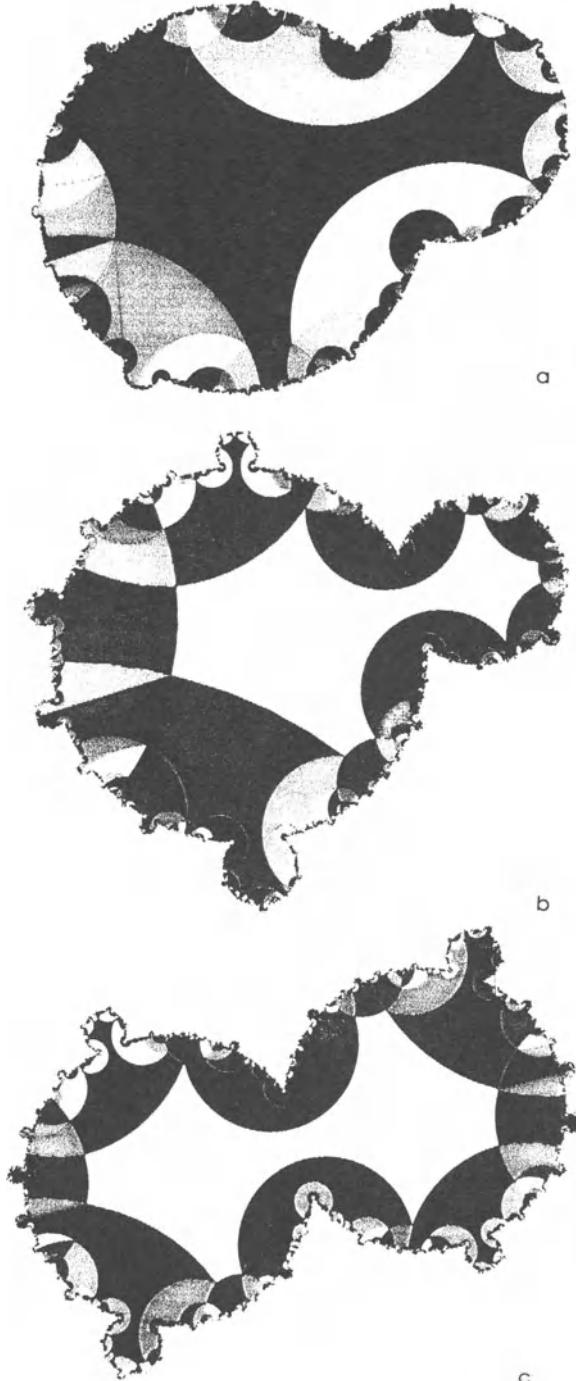


Figure 7a-c. Different realizations of the pattern of base circles give rise to globally non-isometric manifolds. Depicted is a sequence of deformations of the polyhedron F in Fig. 6b. To obtain the polyhedral tiling of H^3 we extend the circular arcs that bound the tiles to circles, and place hemispheres onto them. The fractal boundary curves of the depicted tilings are continuous images of the limit circle in Fig. 6b. They have a Hausdorff dimension $1 < \delta < 2$, but they are not self-similar. These fractal limit sets at the boundary of the covering space

and energy derived from them, which make statements about the asymptotic form of the expansion factor. The predictions derived in this way are fairly unacceptable, despite of attempts to rescue them by means of information theoretical and biological arguments [20]. Dyson claims that time is without end, which naturally raises the question if time is also without beginning...

5. EXTENDED RW-COSMOLOGIES

All that what has been said so far refers to a static 3-space geometry, the spacelike slices (F, Γ) are time independent, and so is the RW-metric in the covering space, apart from the trivial time-dependence via the expansion factor. However, the polyhedron F together with its face-identifying transformations T_i can vary. For example, the α in Fig. 2 may describe some path in the region $|\alpha| > 1$. Different α correspond to globally non-isometric 3-manifolds of curvature -1 , cf. Sec. 2. There is another way to describe this variation of the 3-manifold in its deformation space, the region $|\alpha| > 1$. We keep the polyhedron F as well as the covering group time independent, which is necessary if we want to attach to the 3-manifold a time axis, and vary instead the Poincaré metric g_{ij}^P in H^3 , which is induced onto (F, Γ) . To vary means here that we replace it by a time dependent tensor field \tilde{g}_{ij} , which is periodic, i.e. invariant, with respect to the covering group Γ . Instead of ds^2 we consider $d\tilde{s}^2 = -c^2 d\tau^2 + \tilde{g}_{ij} dx^i dx^j$, $x^i = (x, t)$, on $R^{(+)} \times (F, \Gamma)$. The covering geodesics are now more complicated curves, but the projection mechanism into the 3-manifold remains unchanged.

Let us discuss that a little more explicitly. Imagine that there are two tori, as in Sec. 2, (F, T_α) , $\alpha > 1$, and (F_λ, T_λ) , $T_\lambda : (z, t) \rightarrow \alpha^\lambda (z, t)$, $\lambda > 0$, with F_λ the polyhedron defined by the hemispheres $r = 1$ and $r = \alpha^\lambda$. We construct a diffeomorphism h of H^3 , so that $hT_\alpha h^{-1} = T_\lambda$. We have $h : (x, y, t) \rightarrow (x^2 + y^2 + t^2)^{(\lambda-1)/2}(x, y, t)$, $(x, y, t) := x^i$. If we apply the coordinate transformation h to g_{ij}^P we obtain

$$\tilde{g}_{ij} = a^2(\tau) t^{-2} (\delta_{ij} + (\lambda^2 - 1)(x^2 + y^2 + t^2)^{-1} x^i x^j)$$

This metric is still invariant with respect to the discrete group generated by T_α . If we impose \tilde{g}_{ij} onto the polyhedron F , we get a hyperbolic 3-manifold $(F, T_\alpha, \tilde{g}_{ij})$ which is isometric to (F_λ, T_λ) . This means that we can represent (F_λ, T_λ) by the same covering group and the same polyhedron as the manifold (F, T_α) , provided we replace g_{ij}^P by \tilde{g}_{ij} . Note that the length of the limit cycle of (F_λ, T_λ) is $\lambda \log \alpha$, cf. the caption of Fig. 3. We denote by \tilde{H}^3 the upper half-space endowed with \tilde{g}_{ij} . The 3-manifold represented as $(F, T_\alpha, \tilde{g}_{ij})$ with covering space \tilde{H}^3 we can then easily extend to a 4-manifold using the line element $d\tilde{s}^2$ as indicated above. The crucial point here is that this extension remains possible even for a time-dependent λ . In this case we cannot use (F_λ, T_λ) , $g_{\mu\nu}^{RW}$ (cf. Sec. 4), and $R^{(+)} \times H^3$, because $g_{\mu\nu}^{RW}$ is not invariant with respect to the time-dependent covering group generated by $T_{\lambda(\tau)}$. As pointed out in Sec. 2, $\lambda(\tau)$ describes a path in the deformation space of the topological manifold. The spacelike slices have always curvature $-1/a^2(\tau)$, independent of $\lambda(\tau)$. The energy momentum tensor, defined by the Einstein tensor and ds^2 in $R^{(+)} \times \tilde{H}^3$ projects onto the manifold $R^{(+)} \times (F, T_\alpha, \tilde{g}_{ij})$.

Concerning the technical problem of calculating geodesics and wave fields in the covering space $R^{(+)} \times \tilde{H}^3$ endowed with $d\tilde{s}^2$, we assume that $\lambda(\tau)$ is varying adiabatically. Then ds^2 is approximately generated by applying the transformation

$\tau \rightarrow \tau, x^i \rightarrow h^i(x, \tau)$, to the RW-line element ds^2 , cf. Sec. 4. The $g_{0\mu}$, which make the difference, can be taken into account by Green-Liouville asymptotics. If h^i is independent of time this correspondence between ds^2 and $d\tilde{s}^2$ is exact, and shows how a coordinate transformation in the covering space can give rise to non-isometric manifolds. Infinitesimal gauge transformations are generated by vector fields that are invariant with respect to the covering group, which is clearly not the case with the transformation h^i .

If we consider generic 3-manifolds with non-abelian covering groups, cf. Sec. 3, we have to construct a diffeomorphism h of H^3 , so that the $hT_i h^{-1}$ are again Möbius transformations. This is rather tedious to do in practice, see [5] for a similar problem in two dimensions, and there is a simpler way to realize small global deformations, without explicit knowledge of h .

We construct \tilde{g}_{ij} by adding to the Poincaré metric a field h_{ij}^Γ that is invariant with respect to the covering group Γ , $\tilde{g}_{ij} = a^2(\tau)(t^{-2}\delta_{ij} + h_{ij}^\Gamma)$. Such invariant tensor fields can be generated by periodizing an arbitrary symmetric field $h_{ij}(z, t, \tau)$ on H^3 :

$$h_{ij}^\Gamma(z, t, \tau) = \sum_{\gamma \in \Gamma} h_{kl}(\gamma(x, t), \tau) [\gamma'(z, t)]_i^k [\gamma'(z, t)]_j^l,$$

where γ' denotes the Jacobian of the Möbius transformation. Because \tilde{g}_{ij} is invariant with respect to Γ , $(F, \Gamma, \tilde{g}_{ij}(\tau))$ is a manifold with covering space \tilde{H}^3 . Clearly, for arbitrary fields h_{ij} this manifold will not be of constant curvature. To insure that, the Ricci tensor must be proportional to \tilde{g}_{ij} . If h_{ij}^Γ is small we may linearize $R_{ij}(\tilde{g})$, using as background metric the Poincaré metric. The resulting equation for h_{ij}^Γ is invariant with respect to Γ . Therefore it is enough to find non-periodic solutions h_{ij} and periodize them as shown above, compare the spectral analysis of the electromagnetic field in this context [8].

Finally, I would like to comment on Euclidean cosmologies of finite size, cf. [21], which show a certain resemblance with Mixmaster models, and where it is particularly easy to construct the diffeomorphism h and the deformation space.

We consider the Minkowski line element on $R^{(+)} \times R^3$, and the Euclidean 3-manifold (F_c, Γ_c) , namely the unit cube F_c in R^3 , with opposite faces identified in pairs by Euclidean translations, the generators of Γ_c . Next we consider a parallelepiped \hat{F} generated by the vectors $\vec{a} := (a, 0, 0)$, $\vec{b} := (b_1, b_2, 0)$, $\vec{c} := (c_1, c_2, c_3)$; where a , b_i , c_i are real and opposite faces again identified by Euclidean translations, which generate a group $\hat{\Gamma}$. The Euclidean manifolds (F_c, Γ_c) and $(\hat{F}, \hat{\Gamma})$ are isometric iff F_c and \hat{F} are congruent. We construct readily a linear transformation h in R^3 , that maps F_c onto \hat{F} , so that we have $h\Gamma_c h^{-1} = \hat{\Gamma}$. Evidently $(\hat{F}, \hat{\Gamma})$ is isometric to $(F_c, \Gamma_c, \tilde{g}_{ij})$, $\tilde{g}_{ij} := \delta_{mn} h_{ij}^m h_{ij}^n$. The coefficients a , b_i , c_i are arbitrary functions of time. They parametrize completely the six-dimensional deformation space. If we replace δ_{ij} in the Minkowski metric by \tilde{g}_{ij} we get the line element $d\tilde{s}^2$ of the covering space $R^{(+)} \times R^3$. If $b_1 = c_1 = c_2 = 0$, we have $\tilde{g}_{ij} = \text{diag}[a^2(\tau), b_2^2(\tau), c_3^2(\tau)]$ on the spacelike slices (F_c, Γ_c) . The expansion factors appear here as a path in the deformation space.

6. TOPOLOGICAL BACKSCATTERING AND PARTICLE PRODUCTION IN ELECTROMAGNETIC AND NEUTRINO FIELDS.

In Secs. 2 and 5 we discussed how an open hyperbolic 3-manifold can undergo global metrical deformations, without changing its constant curvature. In fact this condition

can be relaxed, there is no necessity to choose a priori constant curvature, nearly constant curvature and a metric that is uniformly close to the Poincaré metric will likewise do. The covering space formalism is sufficiently flexible to incorporate also large local perturbations and singularities.

In simply connected RW-cosmologies it is known for a long time [22, 23] that variations of the expansion factor can lead to particle creation in quantum fields, and to backscattering in classical fields. I mention here, that one does not need a rapidly varying expansion factor to obtain this effect, linear expansion causes in a simply connected open RW-cosmology particle creation in a massive Dirac field [9].

These effects cannot happen in conformally coupled fields, like neutrinos and electromagnetic fields, because in the solutions of the corresponding wave equations the expansion factor scales out with a simple power law [1, 24]. But global metrical deformations of the 3-space manifold do create particles even in conformally coupled fields, Fig. 8 illustrates how this comes about. We divide the time axis τ for simplicity into

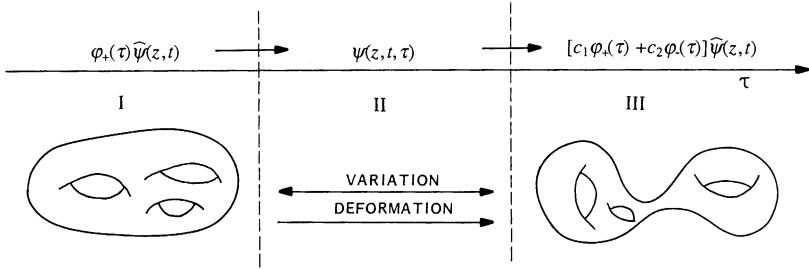


Figure 8. Global metrical deformations of the spacelike slices lead to particle creation and backscattering. A wave packet, initially composed of positive frequencies, receives an admixture of negative frequencies during the deformation.

three intervals. In I and III the metric ds^2 on the spacelike slices is time independent, apart perhaps from a time dependence via the expansion factor, that is irrelevant here. The polyhedron F as well as the covering group is time independent in all three intervals. Under these conditions the wave equation is time separable, and because the field is also conformally coupled, positive and negative frequencies can be unambiguously defined.

In the second interval a global metrical deformation takes place, the line element on the covering space is given now by $d\tilde{s}^2$ in Sec. 5. In this period the wave equation is not separable, and the particle/antiparticle concept loses its meaning here. Finally, in interval III, the general solution of the wave equation is a linear combination of positive and negative frequency modes. If the variation in II is generic, none of the coefficients c_1, c_2 will be zero. Therefore antiparticles are created in the quantum case, and a backscattered wave appears in the case of an electromagnetic field.

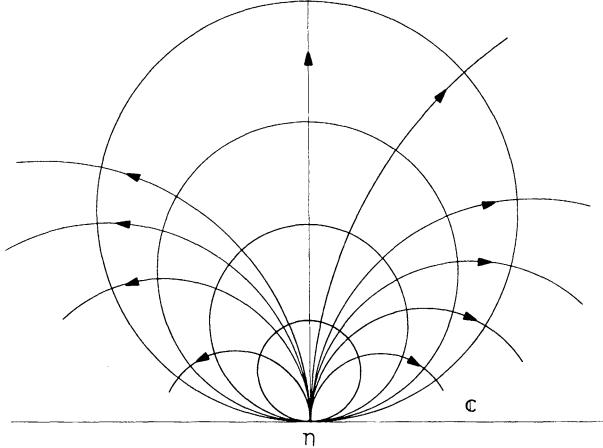


Figure 9. Wave fronts of horospherical elementary waves generated at a point η at infinity of H^3 (the horizon). They appear as the surfaces of constant phase in the eigenfunctions of the Maxwell equations. The bundle of rays issuing from η comprises their orthogonals. Accordingly the surfaces of constant action and phase coincide for rays and elementary waves emanating from a point at infinity.

7. ANGULAR FLUCTUATIONS IN THE TEMPERATURE OF THE MICROWAVE BACKGROUND

Angular anisotropy in the Planck distribution is a possible consequence of global metrical deformations of the spacelike slices. Let us start with a simply connected RW-cosmology and the line element ds^2 in Sec. 4. Because H^3 is homogeneous, it happens that the eikonal of geometric optics appears in the phase of the spectral elementary waves of the Maxwell equations in the covering space. This relation is in other systems only true approximately, semiclassically, here and in homogeneous spaces in general it is exact, cf. Fig. 9.

Geometric optics does not know the concept of momentum. However, we can attach to the rays a momentum via the Einstein relation $p_\mu = \hbar k_\mu$. So we obtain a vector field on H^3 , $p_\mu(z, t, \tau; \eta, s)$, describing the momentum of a horospherical bundle of classical flow lines where s is a spectral parameter, labeling the elementary waves. The directions are labeled by η , cf. Fig. 9. It is easy to see that this flow, coming onto an observer at (z, t) , is isotropic.

We project the horospherical bundles together with the vector fields attached to them into the 3-manifold (F, Γ) . The spectral parameter η is now restricted to the region outside the base circles of the polyhedral faces, f_1 and f_2 in Fig. 6a, namely to the boundary at infinity of the 3-manifold. It is easy to see that for any given $p_\mu(z, t)$ in F there is exactly one ray coming from the boundary, that has this momentum at (z, t) . Thus the flow stays isotropic, and we have again the same Planck distribution $\rho(h\nu/kT)$ as in Minkowski space.

Let us finally switch on adiabatically a global deformation \tilde{g}_{ij} of the 3-space met-

ric, cf. Sec. 5, compare also [25] for the case of local fluctuations. The perturbed horospherical eikonal for rays issuing at some point η on the horizon, cf. Fig. 9, is $\tilde{\psi}^\Gamma = \psi^\Gamma + \chi(z, t, \tau, \eta)$, where χ is a periodic scalar field with respect to the covering group. The perturbed frequencies are $\tilde{\nu} = \nu \left(1 + \frac{1}{\omega} \frac{\partial \chi}{\partial \tau}\right)$, which means that we have to replace in $\rho(h\nu/kT)d\nu$ the temperature by $\tilde{T} = T \left(1 - \frac{1}{\omega} \frac{\partial \chi}{\partial \tau}\right)$, which amounts to an angular variation η of the temperature in the distribution, that remains otherwise unchanged [8]. The adiabatic time dependence of the temperature reminds us that we have here only a first approximation to a non-equilibrium process [26]. A further nice feature of these infinite cosmological models is that we can carry out the thermodynamic limit in them.

8. PARITY VIOLATION DUE TO SELF-INTERFERENCE OF SPACE-REFLECTED WAVES

A space reflection in the covering space H^3 is realized as $P(z, t) = (|z|^2 + t^2)^{-1}(-z, t)$. The center of this reflection is $(0, 1)$; any other point in H^3 will likewise do, but the formula for P gets more complicated. The geometric meaning of $P(z, t)$ is that the point $(0, 1)$ lies always in the middle of the geodesic joining (z, t) and $P(z, t)$.

The universal covering space construction provides here again an easy and explicit way to define a space reflection on the multiply connected 3-space [9]. At first we define the covering projection π^Γ , $H^3 \rightarrow F$, $\pi^\Gamma(z, t) = \gamma^{-1}(z, t)$, if $(z, t) \in \gamma(F)$. We refer here to the tiling of H^3 by images $\gamma(F)$, $\gamma \in \Gamma$, of the fundamental polyhedron F . The space reflection in (F, Γ) we define finally by $P^\Gamma(z, t) = \pi^\Gamma(P(z, t))$. The point of inversion is $\pi^\Gamma(0, 1)$, which lies in the middle of a geodesic joining (z, t) and $P(z, t)$.

The classical geodesic equations are reflection invariant, but the situation is quite different concerning wave mechanics. Imagine a wave packet concentrated on a finite domain in the manifold (F, Γ) . If we apply P^Γ , it can happen that the reflected wave wraps around a handle of the manifold, cf. Fig. 10. Note that the length of the geodesic loop in Fig. 3 can be arbitrarily close to zero. The wave packet starts to interfere with itself, and its norm is not preserved.

Because of this self-interference the space reflection symmetry is already broken on the level of the free Dirac equation. The T -symmetry is likewise broken because of the time dependence of the metric. C is still a good symmetry, but all combinations of C , P , and T are broken. That is remarkable, because usually one has to attach on purpose symmetry breaking interactions in order to achieve this. One can speculate if the baryon asymmetry can be topologically generated. Likewise, if one believes in topologically generated elementary particles, particles as topological/metrical excitations, one could try to understand CP violation as a self-interference effect.

9. CONCLUDING REMARKS

The ultimate aim of cosmology is perhaps to relate the microscopic laws of nature, that we describe now by Newtonian and quantum mechanics, to the global structure and the evolution of the universe. Examples for that are Mach's principle or Dirac's large number hypothesis.

The most important lesson that we have learnt from the revival of classical mechanics during the last two decades is that the instability and unpredictability of classical systems are more the rule than the exception, contrary to that what hitherto has been taught. Dispersion in low dimensional classical systems makes it in practice impossible to view Newton's equations as an initial value problem [5]. The global dynamics in the cosmology presented here is a good example for that. In fact, trajectories which loop a long time close to the chaotic center of the manifold are highly sensitive with respect to the initial data: if the covering trajectory is close to the limit set, then its projection into F consists of many arc pieces, and the initial error multiplies whenever two pieces are glued together.

This instability strongly suggests to use probability densities rather than world lines to describe the classical dynamics adequately. A formalism to do that by means of horospherical flows can be found in [7, 16]. One can then also use the fact that

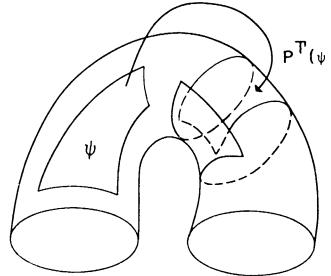


Figure 10. In two dimensions: a wave packet concentrated on a strip. The space reflection wraps the strip around the handle, so that it overlaps with itself. That results in self-interference.

the horospherical eikonal appears in the phase of the quantum mechanical elementary waves, to compare classical with quantum dispersion. Dispersion is inherent in the dynamics, whatever description one uses.

The phenomena reviewed here indicate that from the investigation of the topological structure of space and time further surprising consequences can be expected.

REFERENCES

- [1] Infeld, L. and Schild, A., (1945), *Phys. Rev.*, **68**, 250.
- [2] Schrödinger, E., (1956), *Expanding Universe*, Camb. Univ. Press, Cambridge.
- [3] Jordan, P., (1955), *Schwerkraft und Weltall*, 2nd ed., p. 113, Vieweg, Braunschweig.
- [4] Tomaschitz, R., (1991), *J. Math. Phys.*, **32**, 2571.
- [5] Tomaschitz, R., (1992), *Complex Systems*, **6**, 137.

- [6] Tomaschitz, R., (1993), in *Proceedings of the XIX International Colloquium on Group Theoretical Methods in Physics*, (J. Mateos, ed.), CIEMAT, Madrid.
- [7] Tomaschitz, R., (1993), *J. Math. Phys.*, **34**, 1022.
- [8] Tomaschitz, R., (1993), *J. Math. Phys.*, **34**, 3133.
- [9] Tomaschitz, R., (1993), "Cosmological CP violation", preprint.
- [10] Wheeler, J. A., (1973), in *The Physicist's Conception of Nature*, (J. Mehra, ed.), D. Reidel, Dordrecht.
- [11] Poincaré, H., (1985), *Papers on Fuchsian Functions* (J. Stillwell, ed.), Springer, New York.
- [12] Fenchel, W., (1989), *Elementary Geometry in Hyperbolic Space*, de Gruyter, Berlin.
- [13] Marden, A., (1980), *Bull. Am. Math. Soc. (New Series)* **3**, 1001.
- [14] Maskit, B., (1986), *Kleinian Groups*, Springer, Berlin.
- [15] Tomaschitz, R., (1992), in *Chaotic Dynamics: Theory and Practice*, (T. Bountis, ed.), Plenum, New York.
- [16] Tomaschitz, R., (1994), to appear in *Intern. J. Theoret. Phys..*
- [17] Tomaschitz, R., (1994), in *Fractals in the Natural and Applied Sciences*, (M. Novak, ed.), to appear (Elsevier, Amsterdam).
- [18] Akaza, T., (1964), *Nagoya Math. J.*, **24**, 43.
- [19] Tomaschitz, R., (1992), in *Quantum Chaos-Quantum Measurement*, (P. Cvitanovic, ed.), Kluwer, Dordrecht.
- [20] Dyson, F. J., (1979), *Rev. Mod. Phys.*, **51**, 447.
- [21] Misner, C. W., Thorne, K. S. and Wheeler, J. A., (1973), *Gravitation*, Freeman, San Francisco.
- [22] Schrödinger, E., (1939), *Physica*, **6**, 899.
- [23] Birrell, N. D. and Davies, D. C. W., (1982), *Quantum Fields in Curved Space*, Camb. Univ. Press, Cambridge, UK.
- [24] Parker, L., (1972), *Phys. Rev.*, **D5**, 2905.
- [25] Sachs R. K. and Wolfe, A. M., (1967), *Astrophys. J.*, **147**, 73.
- [26] Israel, W., (1972), in *General Relativity*, (L. O'Raifeartaigh, ed.), Clarendon Press, Oxford.

RELATIVISTIC FRACTAL COSMOLOGIES

Marcelo B. Ribeiro¹

Departamento de Astrofísica, Observatório Nacional-CNPq,
Rua General José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil

Abstract. This article presents a review of an approach for constructing a simple relativistic fractal cosmology, whose main aim is to model the observed inhomogeneities of the distribution of galaxies by means of the Tolman solution of Einstein's field equations for spherically symmetric dust in comoving coordinates. Such model is based on earlier works developed by L. Pietronero and J. R. Wertz on Newtonian cosmology, and the main points of these models are also discussed. Observational relations in Tolman's spacetime are presented, together with a strategy for finding numerical solutions which approximate an averaged and smoothed out single fractal structure in the past light cone. Such fractal solutions are actually obtained and one of them is found to be in agreement with basic observational constraints, namely the linearity of the redshift-distance relation for $z < 1$, the decay of the average density with the distance as a power law (the de Vaucouleurs' density power law), the fractal dimension within the range $1 \leq D \leq 2$, and the present range of uncertainty for the Hubble constant. The spatially homogeneous Friedmann model is discussed as a special case of the Tolman solution, and it is found that once we apply the observational relations developed for the fractal model we find that all Friedmann models look inhomogeneous along the backward null cone, with a departure from the observable homogeneous region at relatively close ranges. It is also shown that with these same observational relations the Einstein-de Sitter model can have an interpretation where it has zero global density, a result consistent with the "zero global density postulate" advanced by Wertz for hierarchical cosmologies and conjectured by Pietronero for fractal cosmological models. The article ends with a brief discussion on the possible link between this model and nonlinear and chaotic dynamics.

¹ e-mail: mbr@on.br

1. INTRODUCTION

In the study of dynamical systems it is usual to approach the problem under consideration by following a more or less pre-determined path, which can be sketched as being basically made of four main stages. In the FIRST step one usually sets up the differential or difference equations that describe the system in study, that is, one defines the dynamical system itself. The SECOND step would be the determination of orbits of the dynamical system, their discretization, that is, determining the Poincaré section, etc, while the THIRD stage usually consists of attempting to determine whether or not the system exhibits sensitivity to initial conditions, and this could be done by studying the Lyapunov exponents of the orbits. In other words, at this stage one usually seeks to determine whether or not the dynamical system is chaotic. Then the FOURTH stage would be the search for fractal self-similarity and strange attractors, with the possible determination of the fractal dimension of the self-similar pattern.

I will call this way of studying dynamical systems as “the mathematicians’ approach” as it is usually followed by them, and the point I would like to emphasize is that in such an approach the self-similarity exhibited by the system, and the determination of its fractal dimension, will play a minor and side role, and may even be expendable in the whole treatment since when one finally reaches that stage most of the dynamical behaviour of the system would already be elucidated.

However, when a physicist looks at his or her noisy data, or in a cosmological context, at his or her messy observations, the physicist will usually attempt to get some sense from his/her real data, and in this process he or she may ask: is this noisy data really noisy? Could some sort of pattern be identified in this data? Then, the physicist may go even further and pose the following question. Could a chaotic behaviour be somehow represented in this data?

To attempt to answer those questions, in particular the last one, the physicist might wish to approach the problem from the opposite direction taken by the mathematicians, and a path that could be taken may be sketched as follows. The FIRST stage would be to use the self-similarity exhibited by the system in order to determine its fractal dimension, together with some minimal and basic dynamical assumptions about the system such that one starts with a workable and testable model. Then the SECOND step would be to try to determine orbits, compute the Lyapunov exponents, make the Poincaré map, try to find the attractors, etc, of the system in order to reach the THIRD stage which would be to ascertain whether or not the system exhibits chaos. The FOURTH stage would consist in the determination of the physical processes and their observable effects that such possible chaotic behaviour would bring to the system under study.

It is therefore clear that in this “physicist’s approach” fractals would play an essential and primary role in modelling natural phenomena as the physicist would start by attempting to model structures we can see, assuming implicitly what is in essence Mandelbrot’s view of fractal geometry [1] where one seeks to model naturally occurring shapes like mountains, clouds and coastlines. However, in order to follow the path outlined above, the physicist needs first to do two things:

1. recognize a fractal pattern in the data;
2. characterize it in some way.

Unfortunately, these two points lead to questions that turn out to be easy to ask but hard to answer, and in a cosmological context both points above are currently controversial, especially the first one, inasmuch as those who voiced their recognition of a fractal pattern in the large-scale distribution of galaxies have faced stiff resistance and, not rarely, hostility.

This article intends to present a review of a specific approach to the use of fractals in relativistic cosmology inspired in Luciano Pietronero's Newtonian model [2], although most of the treatment was first developed by James R. Wertz in his PhD thesis [3], where he studied Newtonian Hierarchical Cosmology, and rediscovered by Pietronero by means of the fractal language, as we shall see next. The underlining philosophy of this approach is, as discussed above, to take advantage of the ability of fractals in modelling shapes, and the basic aim is to find solutions of Einstein's field equations with an approximate *single* fractal behaviour along the past light cone. Therefore, multifractals are not considered here. This is basically a descriptive model which resembles Mandelbrot's characterization of coastlines as fractal shapes [1], and so far only covers the first stage of the physicist's approach to the study of dynamical systems as described above.

2. MODERN COSMOLOGY AND FRACTALS

The goal of modern cosmology is to find the large-scale matter distribution and spacetime structure of the universe from astronomical observations and, broadly speaking, there are two different ways of approaching this problem [4]. The most popular and standard approach consists of some *a priori* assumptions about the geometry of the universe, usually based on pragmatic and philosophical reasons. This generally reduces the cosmological problem to the determination of a few free parameters that characterize such universe models, and this determination then becomes the primary objective of observational cosmology. That standard approach assumes the *cosmological principle*, where the universe is thought to be spatially homogeneous and isotropic on large-scales, and is represented by the maximally symmetric Friedmann spacetime.

The alternative approach of studying cosmology is to attempt as far as possible to determine spacetime geometry directly from astronomical observations, where any kind of *a priori* assumptions are kept to a minimum dictated by the essential and basic requirements necessary such that one is able to construct a workable model.

The relativistic fractal cosmology of this article approaches the cosmological problem from this alternative point of view, although it must be said, it is obviously just one possible way of looking at this problem. Considering recent astronomical observations as fundamental empirical facts, I shall attempt to "guess" a metric in a specific form suggested by those same observations which enables us to model a more real, lumpy universe as detected in our past light cone.

The fundamental empirical facts of this approach to cosmology are the recent all-sky redshift surveys [5, 6, 7, 8, 9, 10] where it is clear that the large-scale structure of the universe does not show itself as a smooth and homogeneous distribution of luminous matter as was thought earlier. Rather the opposite, since up to the limits of the observations presented in those surveys, the three-dimensional cone maps show a very inhomogeneous picture, with galaxies mainly grouped in clusters or groups alongside regions devoid of galaxies, virtually empty spaces with scales of the same order of magnitude as their neighbour clusters. Such mapping of the skies gives us a picture of

the distribution of galaxies as a complex mixture of interconnected voids, clusters and superclusters.

From these observations, what is obvious for the eyes is the pattern that appears to be common in all surveys: the deeper the probing is made, the more similar structures are observed and mapped, with clusters turning into superclusters and even bigger voids being identified. As the size of these clusters are only limited by the depth of the surveys themselves, there is so far no empirical evidence of where and if those structures finish.

With respect to this pattern, two ideas seem to fit in. The first is the old concept of *hierarchical clustering* [11, 12] which states that galaxies join together to form clusters that form superclusters which themselves are elements of super-superclusters and so on, possibly ad infinitum. The second is the more recent concept of *fractals*, or self-similar structures, of which a rather loose tentative definition proposed by Mandelbrot seems to be adequate for the present purpose: “a fractal is a shape made of parts similar to the whole in some way” (see [13], p. 11). In this context self-similarity means that a fractal consists of a system in which more and more structure appears at smaller and smaller scales and the structure at small scales is similar to the one at large scales. In other words, the same structure repeats itself at different scales. It is therefore clear that fractals are a more precise version of the same *scaling* idea behind the concept of hierarchical clustering, where clusters and superclusters form a self-similar pattern repeating itself at different scales. Consequently, in this context to talk about a hierarchical structure is basically the same as to talk about a fractal system.

The attempt of modelling the large-scale structure of the universe as a hierarchical or fractal structure is not at all a new idea, and in order to give some historical perspective of this problem I shall present below a historical summary of the main contributions of such attempts.²

1907 E. E. Fournier D’Albe published the proposal of a hierarchical construction of the universe [14];

1908, 1922 C. V. L. Charlier applied Fournier D’Albe’s idea to an astronomical world model in order to explain the Olber’s paradox [11, 12];

1922-1924 F. Selety and A. Einstein discussed “Charlier’s hierarchical model” [15, 16, 17, 18].

1970 G. de Vaucouleurs resurrected Charlier’s model [19, 20] by proposing a hierarchical cosmology as a way of explaining his finding that galaxies seem to follow an average density power law with negative slope, of the form $\bar{\rho} \propto r^{-1.7}$; J. R. Wertz studied some possible models of a Newtonian hierarchical cosmology [3, 21], and his simplified “regular polka-dot model” anticipated Pietronero’s [2] fractal treatment;

1971-1972 M. J. Haggerty and J. R. Wertz developed further the Newtonian hierarchical models and proposed some possible observational tests [22, 23]. The debate

² The history of the hierarchical clustering hypothesis is of being forgotten by most every time it is voiced by some as being a natural way of constructing the universe, only to be resurrected by somebody else, just a while later, who not rarely is partially or completely unaware of many of the previous results and studies. Therefore, it is quite reasonable to believe that this historical summary may still be revised once possible isolated and forgotten works are unearthed.

concerning the motivations [24] and observational feasibility of hierarchical cosmological models is re-initiated and centered around the possible deviations from local expansion [23, 25]. Such debate continued for over a decade or so, and now it seems to favour the view that there really are such deviations (see [26, 27] and references therein);

- 1972** W. B. Bonnor studied [28] a relativistic model in a Tolman spacetime with the de Vaucouleurs' density power law $\bar{\rho} \propto r^{-1.7}$;
- 1978-1979** P. S. Wesson studied relativistic models [29, 30] with homothetic self-similarity where the local density followed $\rho \propto r^{-2}$ at some epochs;
- 1987** L. Pietronero published his fractal model [2] where the de Vaucouleurs density power law $\bar{\rho} \propto r^{-\gamma}$ is obtained once one assumes that the large-scale distribution of galaxies forms a self-similar single fractal system. He also voiced sharp criticisms of the use of the spatial two-point correlation function for the characterization of the distribution of galaxies, in particular its indication that a homogenization of the distribution occurs at about 5 Mpc, a result he called "spurious" and due to the widespread unquestioned, and often unjustified, assumption of the untested homogeneous hypothesis. The debate still rages on [31, 32, 33, 34, 35, 36];
- 1988** R. Ruffini, D. J. Song & S. Taraglio proposed that the fractal system should have an upper cutoff to homogeneity in order to solve what they believed was an apparent conflict between "the commonly accepted idea in theoretical cosmology that greater distances represent earlier epochs of the Universe implying that higher average densities should be observed", and the de Vaucouleurs density power law that implies the opposite [37];³ D. Calzetti, M. Giavalisco & R. Ruffini studied the spatial two-point correlation function for galaxies under a fractal perspective [38] and reached some conclusions similar to Pietronero's [2].

Since 1988 there has been a flurry of activity on fractal cosmology, and the different approaches to the problem are in many ways opposing to each other, a fact which has obviously created a lot of debate about the issue. Nevertheless, from this historical summary it is clear that although hierarchical cosmology has never belonged to the mainstream of research in cosmology (at least until recently), it did attract the attention of many researchers, who seriously considered the observed scaling behaviour of the distribution of galaxies as something fundamental in cosmology.

Despite such interest, the most recent pre-fractal hierarchical cosmological models suffered from some important shortcomings which may explain why their development was hampered. In the first place it was not generally clear what one meant mathematically by the idea of "galaxies joining to form clusters that form superclusters which themselves are elements of superclusters, and so on", and such lack of clarity had the effect of making the hierarchical concept not well defined mathematically.⁴ From a fractal perspective, it is obvious that what is behind the hierarchical clustering concept is the scaling behaviour empirically accepted by many in the distribution of galaxies, and hence the main descriptor of such behaviour becomes the fractal dimension which

³ Actually, such a possible transition to homogeneity had already been suggested by Wertz, [3] p. 27.

⁴ It should be noted here that Wertz [3, 21] did discuss the hierarchical concept in greater mathematical details, but that work seems to have attracted little attention since it has been published.

must be appropriately defined in the context of the distribution of galaxies. Therefore, it is the exponent of the de Vaucouleurs density power law⁵ that has a fundamental physical meaning, and not the law itself which is a consequence of the self-similarity of the distribution of galaxies [39]. However, the hierarchical clustering concept had to wait until the appearance of fractals and the works of Mandelbrot [1] and Pietronero [2], who clarified this point.

A second difficulty of pre-fractal hierarchical cosmology was their relativistic models which have never expressed distance by means of observables, always using the unobservable coordinate distance r (in the spherically symmetric models) as measure of distance. That created an additional difficulty between the empirically observed hierarchical concept and the relativistic models, as the latter became in fact distorted attempts of modelling the de Vaucouleurs' density power law since the relativistic hierarchy was unobservable. Wesson [29, 30] even gave up using the average density $\bar{\rho}$ and replaced it by the local density ρ in his version of de Vaucouleur's density power law, and this inevitably led him to produce a model where the essence of the original hierarchical clustering hypothesis was lost. On top of all that, none of the relativistic models evaluated densities along the past null geodesic despite the fact that General Relativity states this is where astronomical observations are actually made. Consequently, the pre-fractal relativistic hierarchical cosmologies were in fact ill represented attempts of modelling hierarchical clustering, caused mainly by the usual difficulties of modelling physical phenomena by means of General Relativity conjugated with the additional difficulties of solving the past null geodesic equation. Nevertheless, despite their shortcomings those initial models did put forward interesting ideas and methods, and as we shall see next the present approach to relativistic fractal cosmology borrows some ideas from those previous relativistic models and attempts to address and overcome most of these problems, with the aim of recapturing the empirical and scaling essence of the de Vaucouleurs density power law.

3. PIETRONERO AND WERTZ'S MODELS

Before we discuss the relativistic fractal models themselves, let us briefly see the main most basic points of both Pietronero and Wertz's models and how they compare to each other. While the former model is nowadays relatively well-known in the literature, the latter seems to have attracted very little attention since its publication as it is very rarely quoted. For this reason I shall produce a more detailed presentation of the basis of Wertz's model such that one may be able to fully appreciate its importance.

In developing his model, Pietronero [2] started with a schematic representation of a deterministic fractal structure, reproduced in figure 1, whose basic idea is as follows. If we start from a point occupied by an object and count how many objects are present within a volume characterized by a certain length scale, we get N_0 objects within a radius r_0 , $N_1 = \tilde{k}N_0$ objects within a radius $r_1 = kr_0$, $N_2 = \tilde{k}N_1 = \tilde{k}^2N_0$ objects within a radius $r_2 = kr_1 = k^2r_0$. In general we have

$$N_n = \tilde{k}^n N_0 \quad (1)$$

⁵ From now on I shall call by "de Vaucouleurs density power law" the decay of the average density of the distribution of galaxies with the distance (using whatever definition of distance), and given generically by the form $\bar{\rho} \propto r^{-\gamma}$, where $0 \leq \gamma < 3$, that is, γ does not necessarily have to be the specific value of 1.7 as originally found by de Vaucouleurs [19].

within

$$r_n = k^n r_0, \quad (2)$$

where \tilde{k} and k are constants. By taking the logarithm of equations (1) and (2) and

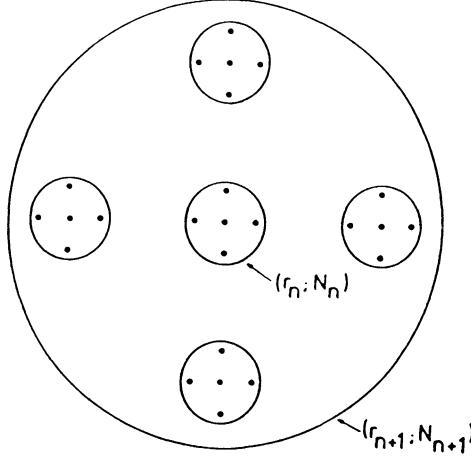


Figure 1. Reproduction from [2] of a schematic illustration of a deterministic fractal system from where a fractal dimension can be derived. The structure is self-similar, repeating itself at different scales.

dividing one by the other we get

$$N_n = \sigma(r_n)^D, \quad (3)$$

with

$$\sigma \equiv \frac{N_0}{(r_0)^D}, \quad (4)$$

$$D \equiv \frac{\log \tilde{k}}{\log k}, \quad (5)$$

where σ is a prefactor of proportionality related to the lower cutoffs N_0 and r_0 of the fractal system, that is, the inner limit where the fractal systems ends, and D is the fractal dimension. If we smooth out the fractal structure we get the continuum limit of equation (3),

$$N(r) = \sigma r^D, \quad (6)$$

called “generalized mass-length relation” by Pietronero. The de Vaucouleurs density power law is obtained if we now suppose that a portion of the fractal system is contained inside a spherical sample of radius R_s . Then

$$\langle \rho \rangle \equiv \frac{N(R_s)}{V(R_s)} = \frac{3\sigma}{4\pi} (R_s)^{-\gamma}; \quad \gamma \equiv 3 - D, \quad (7)$$

where $\langle \rho \rangle$ is the average density of the distribution. If we take the value of γ as the one found by de Vaucouleurs [19], this implies that the fractal dimension of the distribution is $D \approx 1.3$.

These are the results of interest for the relativistic approach of this article, and therefore I shall stop here this brief summary of Pietronero's fractal model. The interested reader can find in [35] a comprehensive account of this fractal approach to cosmology, plus the controversy surrounding the spatial and angular two-point correlation functions and a discussion on multifractals in this context.

The hierarchical model advanced by Wertz [3] was conceived at a time when fractal ideas had not yet appeared, so in developing his model Wertz was forced to start with a more conceptual discussion in order to offer "a clarification of what is meant by the 'undefined notions' which are the basis of any theory" ([3], p. 3). Then he stated that "a cluster consists of an aggregate or gathering of ELEMENTS into a more or less well-defined group which can be to some extent distinguished from its surroundings by its greater density of elements. A HIERARCHICAL structure exists when iTH ORDER CLUSTERS are themselves elements of an (i+1)TH ORDER CLUSTER. Thus, galaxies (ZERO TH ORDER CLUSTERS) are grouped into FIRST ORDER CLUSTER. First order clusters are themselves grouped together to form SECOND ORDER CLUSTERS, etc, *ad infinitum*" (see figure 2).

Although this sort of discussion may be very well to start with, it demands a precise definition of what one means by a cluster in order to put those ideas on a more solid footing, otherwise the hierarchical structure one is talking about continues to be a somewhat vague notion. Wertz seemed to have realized this difficulty when later he added that "to say what percentage of galaxies occur in clusters is beyond the abilities of current observations and involves the rather arbitrary judgment of what sort of grouping is to be called a cluster. (...) It should be pointed out that there is not a clear delineation between clusters and superclusters" (p. 8).

Despite this initially descriptive and somewhat vague discussion about hierarchical structure, which is basically a discussion about scaling in the fractal sense, Wertz did develop some more precise notions when he began to discuss specific models for hierarchy, and his starting point was to assume what he called the "universal density-radius relation", that is, the de Vaucouleurs density power law, as a fundamental empirical fact to be taken into account in order to develop a hierarchical cosmology. Then if $M(x, r)$ is the total mass within a sphere of radius r centered on the point x , he defined the *volume density* ρ_v as being the average over a sphere of a given volume containing M . Thus

$$\rho_v(x, r) \equiv \frac{3M(x, r)}{4\pi r^3}, \quad (8)$$

and the *global density* was defined as being

$$\rho_g \equiv \lim_{r \rightarrow \infty} \rho_v(x, r). \quad (9)$$

A *pure hierarchy* is defined as a model universe which meets the following postulates:
(i) for any positive value of r in a bounded region, the volume density has a maximum;
(ii) the model is composed of only mass points with finite non-zero mean mass;
(iii) the zero global density postulate: "for a pure hierarchy the global density exists and is zero everywhere" (see [3] p. 18).

With this picture in mind, Wertz states that "in any model which involves clustering, there may or may not appear discrete lengths which represent clustering on

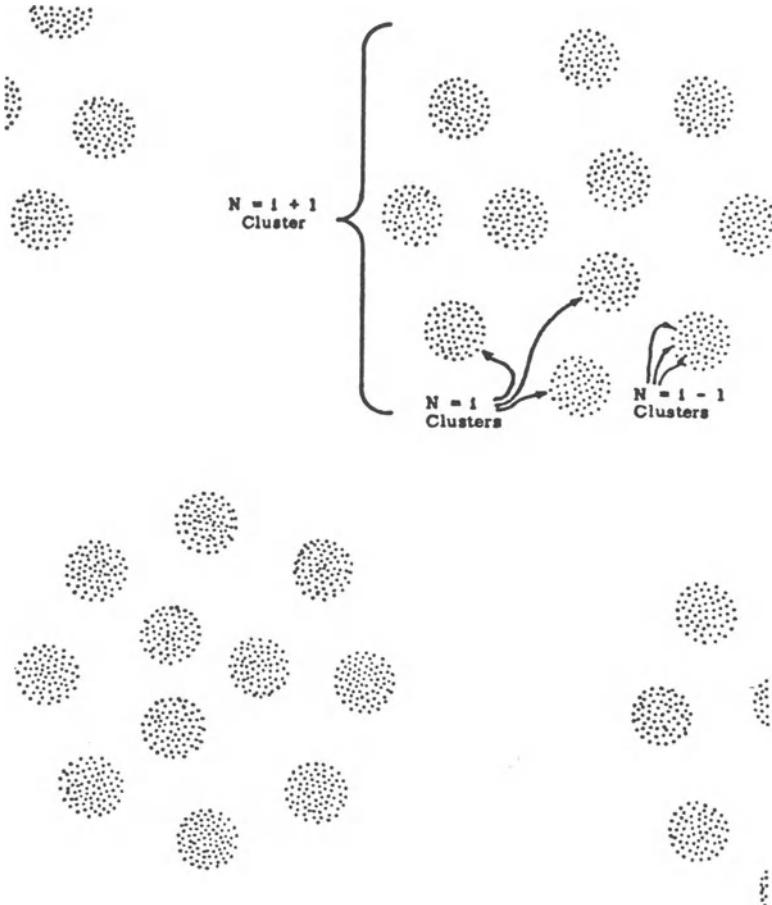


Figure 2. Reproduction from p. 25 of [3] of a rough sketch cross-section of a portion of an $N = i + 2$ cluster of a polka dot model.

different scales. If no such scales exist, one would have an INDEFINITE HIERARCHY in which clusters of every size were equally represented (...). At the other extreme is the DISCRETE HIERARCHY in which cluster sizes form a discrete spectrum and the elements of one size cluster are all clusters of the next lowest size" (p. 23). Then in order to describe *polka dot models*, that is, structures in a discrete hierarchy where the elements of a cluster are all of the same mass and are distributed regularly in the sense of crystal lattice points, it becomes necessary for one be able to assign some average properties. So if N is the order of a cluster, $N = i$ is a cluster of arbitrary order (figure 2), and at least in terms of averages a cluster of mass M_i , diameter D_i and composed of n_i elements, each of mass m_i and diameter d_i , has a density given by

$$\rho_i = \frac{6M_i}{\pi D_i^3}. \quad (10)$$

From the definitions of discrete hierarchy it is obvious that

$$M_{i-1} = n_{i-1}m_{i-1} = m_i, \quad (11)$$

and if the ratio of radii of clusters is

$$a_i \equiv \frac{D_i}{d_i} = \frac{D_i}{D_{i-1}}, \quad (12)$$

then the *dilution factor* is defined as

$$\phi_i \equiv \frac{\rho_{i-1}}{\rho_i} = \frac{a_i^3}{n_i} > 1, \quad (13)$$

and the *thinning rate* is given by

$$\theta_i \equiv \frac{\log(\rho_{i-1}/\rho_i)}{\log(D_i/D_{i-1})} = \frac{\log(a_i^3/n_i)}{\log a_i}. \quad (14)$$

A *regular polka dot model* is defined as the one whose number of elements per cluster n_i and the ratio of the radii of successive clusters a_i are both constants and independent of i , that is, n and a respectively. Consequently, the dilution factor and the thinning rate are both constants in those models,

$$\phi = \frac{a^3}{n}, \quad \theta = \frac{\log(a^3/n)}{\log a}. \quad (15)$$

The *continuous representation* of the regular polka dot model, which amounts essentially to writing the hierarchical model as a continuous distribution, is obtained if we consider r , the radius of spheres centered on the origin, as a continuous variable. Then, from equation (12) the radius of the elementary point mass r_0 , is given by

$$r_0 = \frac{R_1}{a}, \quad (16)$$

where R_N is the radius of a N th order cluster with M_N mass, V_N volume, and obviously that $R_0 = r_0$. It follows from equation (16) the relationship between N and r ,

$$r = a^N r_0, \quad (17)$$

where $R_N = r$. Notice that by doing this continuous representation Wertz ended up obtaining an equation (eq. 17) which is nothing more than exactly equation (2) of Pietronero's single fractal model, although Wertz had reached it by means of a more convoluted reasoning. Actually, the critical assumption which makes his polka dot model essentially the same as Pietronero's fractal model was to assume the regularity of the model because then a and n become constants. Also notice that this continuous representation amounts to changing from discrete to an indefinite hierarchy, where in the latter the characteristic length scales for clustering are absent. Therefore, in this representation clusters (and voids) extend to all ranges where the hierarchy is defined with their sizes extending to all scales between the inner and possible outer limits of the hierarchy. Hence, in this sense the continuous representation of the regular polka dot model has exactly the same sort of properties as the fractal model discussed by Pietronero.

From equation (11) we clearly get

$$M_N = n^N M_0, \quad (18)$$

which is equal to equation (1), except for a different notation, and hence the de Vaucouleurs density power law is easily obtained as

$$\rho_v = \frac{M_N}{V_N} = \left[\frac{3M_0}{4\pi r_0^{(\log n/\log a)}} \right] r^{-\theta}, \quad (19)$$

where θ is the thinning rate

$$\theta = 3 - \left(\frac{\log n}{\log a} \right). \quad (20)$$

Notice that equations (19) and (20) are exactly equations (7), where γ is now called the thinning rate. Finally, the *differential density*, called *conditional density* by Pietronero, is defined as

$$\rho_d \equiv \frac{1}{4\pi r^2} \frac{dM(r)}{dr} = \left(1 - \frac{\theta}{3} \right) \rho_v. \quad (21)$$

From the presentation above it is then clear that from a geometrical viewpoint Wertz's continuous representation of the regular polka dot model is nothing more than Pietronero's single fractal model. However, the two approaches may be distinguished from each other by some important conceptual differences. Basically, as Pietronero clearly defines the exponent of equation (3) as a fractal dimension, that immediately links his model to the theory of critical phenomena in physics, and also to nonlinear dynamical systems, bringing a completely new perspective to the study of the distribution of galaxies, with potentially new mathematical concepts and analytical tools to analyze this problem. In addition, it also strongly emphasizes the fundamental importance of scaling behaviour in the observed distribution of galaxies and the exponent of the power law, as well as pointing out the appropriate mathematical tool to describe this distribution, namely the fractal dimension. All that is missing in Wertz's approach, and his thinning rate is just another parameter in his description of hierarchy, without any special physical meaning attached to it. Therefore, in this sense his contribution started and remained as an isolated work, forgotten by most, and which could even be viewed simply as an ingenious way of modelling Charlier's hierarchy, but nothing more.

Nonetheless, it should be said that this discussion must not be viewed as a critique of Wertz's work, but simply as a realization of the fact that at Wertz's time nonlinear dynamics and fractal geometry were not as developed as at Pietronero's time, if developed at all, and therefore Wertz could not have benefited from those ideas. Despite this it is interesting to note that with less data and mathematical concepts he was nevertheless able to go pretty far in discussing scaling behaviour in the distribution of galaxies, developing a model to describe it in the context of Newtonian cosmology, and even suggesting some possible ways of investigating relativistic hierarchical cosmology.

4. A RELATIVISTIC APPROACH TO HIERARCHICAL (FRACTAL) COSMOLOGY

In this section I shall start to develop a relativistic cosmology based on the ideas expressed by Pietronero and Wertz and discussed in the previous section. The notation used will be the same as in Pietronero's approach with minor changes, but some quantities will be referred by the names used by Wertz.

The first thing necessary for one to start a relativistic model is obviously the choice of the appropriate metric for the problem under consideration. In the case of

a relativistic fractal cosmology, we need an inhomogeneous metric so that it becomes possible to derive a relativistic version of Pietronero's relation (6). Following the simple geometrical ideas outlined by both Wertz and Pietronero, spherical symmetry seems to be reasonable enough to start with. By means of a similar reasoning, a dust distribution of matter also seems reasonable enough to start with, and in that way we have outlined some requirements which are equivalent to making some strong simplifications that seem enough for a simple exploratory relativistic fractal cosmological model. Bearing this discussion in mind, the *Tolman solution* suggests itself as it is the general solution of the Einstein's field equations for spherically symmetric dust in comoving coordinates [40]. However, the Tolman solution is spherically symmetric about *one* point, and constructing a cosmological model with it means that we would be giving up the Copernican principle which states that there are no preferred points in the universe. Despite this difficulty, if we assume that there could be an upper cutoff to homogeneity in our fractal system, we can construct a model using a variation of the Einstein-Straus geometry, also known as "Swiss-cheese" models, with an interior solution provided by the Tolman metric surrounded by a Friedmann spacetime. Such a model needs to solve the junction conditions between the two metrics in order to achieve a smooth transition, and the solution imposes strong restrictions to the mass inside the Tolman region, namely that the gravitational mass inside must be the same as if the whole spacetime were Friedmannian and the Tolman region were never there [41]. By means of such scheme we are able to satisfy the Copernican principle in our relativistic fractal model, but for reasons that will become clear later, such geometry is not really mandatory.⁶

In order to make use of Tolman's models as descriptors of observations, it is necessary first of all to derive the observational relations in this metric. I shall present next a brief summary of the Tolman spacetime followed by its observational relations, but without demonstration. Full details about the observational relations in this metric can be found in [41].

The Tolman metric with $\Lambda = 0$ and $c = G = 1$ may be written as

$$dS^2 = dt^2 - \frac{R'^2}{f^2} dr^2 - R^2 d\Omega^2; \quad r \geq 0, \quad R(r, t) > 0, \quad (22)$$

where

$$d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2,$$

and $f(r)$ is an arbitrary function. The Einstein's field equations for the metric (22) reduces to a single equation,

$$2R\dot{R}^2 + 2R(1 - f^2) = F, \quad (23)$$

where $F(r)$ is another arbitrary function. The proper density is given by

$$8\pi\rho = \frac{F'}{2R'R^2}, \quad (24)$$

and the dot means $\partial/\partial t$ and the prime $\partial/\partial r$.

⁶ See [41] for a more detailed discussion of the role played by the Copernican and cosmological principles in fractal cosmologies. It is interesting to mention that Wertz [3] did suggest the Swiss-cheese model as a possible way of modelling relativistic hierarchy, and he also made a discussion about the cosmological principle in the context of hierarchical cosmologies.

The solution of equation (23) has three distinct cases according as $f^2 = 1$, $f^2 > 1$ and $f^2 < 1$, these cases being termed, respectively, parabolic, hyperbolic and elliptic Tolman models. In the parabolic models ($f^2 = 1$) the solution of equation (23) is

$$R = \frac{1}{2}(9F)^{1/3}(t + \beta)^{2/3}, \quad (25)$$

where $\beta(r)$ is a third arbitrary function.⁷ In the hyperbolic models ($f^2 > 1$) the solution of equation (23) may be written in terms of a parameter Θ

$$R = \frac{F(\cosh 2\Theta - 1)}{4(f^2 - 1)}, \quad (26)$$

$$t + \beta = \frac{F(\sinh 2\Theta - 2\Theta)}{4(f^2 - 1)^{3/2}}, \quad (27)$$

and finally in the elliptic models ($f^2 < 1$) the solution of equation (23) may be written as

$$R = \frac{F(1 - \cos 2\Theta)}{4|f^2 - 1|}, \quad (28)$$

where

$$t + \beta = \frac{F(2\Theta - \sin 2\Theta)}{4|f^2 - 1|^{3/2}}. \quad (29)$$

The three classes of the Tolman solution are the equivalent of flat, open and closed Friedmann models.

The observational relations in Tolman's spacetime necessary in this work have been calculated in [41], and I shall only state here the results obtained. The *luminosity distance* in the Tolman model is given by

$$d_L = R(1 + z)^2. \quad (30)$$

The *redshift* may be written as

$$1 + z = (1 - I)^{-1}, \quad (31)$$

where the function $I(r)$ is the solution of the differential equation⁸

$$\frac{dI}{dr} = \frac{\dot{R}'}{f}(1 - I). \quad (32)$$

The number of sources which lie at radial coordinate distances less than r as seen by the observer at $r = 0$ and along the past light cone, that is, the *cumulative number count*, is given by

$$N_c(r) = \frac{1}{4M_G} \int \frac{F'}{f} dr, \quad (33)$$

where M_G is the average galactic rest mass ($\sim 10^{11} M_\odot$). The volume V of the sphere which contains the sources, and the volume (average) density ρ_v , have the form

$$V(r) \equiv \frac{4}{3}\pi(d_L)^3 = \frac{4}{3}\pi R^3(1 + z)^6, \quad (34)$$

⁷ Actually, one of the three functions $f(r)$, $F(r)$, $\beta(r)$ can be removed by a coordinate transformation, so there really are two arbitrary functions in the Tolman solution.

⁸ The expression for the redshift in the Tolman model was obtained in collaboration with M. A. H. MacCallum.

$$\rho_v \equiv \frac{N_c(r)M_G}{V(r)} = \frac{3}{16\pi R^3(1+z)^6} \int \frac{F'}{f} dr. \quad (35)$$

The relativistic version of Pietronero's generalized mass-length relation used in this work is given by

$$N_c = \sigma(d_\ell)^D, \quad (36)$$

and if we substitute equations (34) and (36) into equation (35) we get a relativistic version of the de Vaucouleurs density power law

$$\rho_v = \frac{3\sigma M_G}{4\pi} (d_\ell)^{-\gamma}, \quad (37)$$

where

$$\gamma = 3 - D \quad (38)$$

is Wertz's thinning rate. Equation (37) gives the volume density for an observed sphere of certain radius d_ℓ that contains a portion of the fractal system. All observational relations above must be calculated along the past light cone, so if we adopt the radius coordinate r as the parameter along the backward null cone, we can then write the radial null geodesic of metric (22) as

$$\frac{dt}{dr} = -\frac{R'}{f}. \quad (39)$$

Notice that along the past light cone $R = R[r, t(r)]$, where $t(r)$ is the solution of equation (39), and this is the function to be used in the relations above such that they are really calculated along the null geodesic.

The next step in order to find fractal solutions in the Tolman model is to take advantage of its own freedom, given by the arbitrariness of the functions $f(r)$, $F(r)$ and $\beta(r)$, such that we are able to simulate the desired distribution of dust by means of a method similar to the one employed by Bonnor [28], but in a more complex and realistic context than his. However, in order to do so we have to solve the past radial null geodesic (39) and then the equation for the redshift (32), an almost impossible task to be carried out analytically due to the fact that the functions R , R' , \dot{R}' form complex algebraic expressions [41], and so a numerical solution becomes inevitable. Such an approach can be essentially described as follows.

For a fractal structure as described above, the de Vaucouleurs density power law holds. Hence equation (37) may be written as

$$\log \rho_v = a_1 + a_2 \log d_\ell, \quad (40)$$

where a_1 and a_2 are constants related to the lower cutoff of the fractal pattern and its fractal dimension,

$$D = a_2 + 3, \quad (41)$$

$$\sigma = \frac{4\pi}{3M_G} \exp(a_1). \quad (42)$$

Generally speaking, the numerical algorithm for finding those fractal solutions can be summarized as follows:

1. START by choosing $f(r)$, $F(r)$, $\beta(r)$;
2. solve numerically the two ordinary differential equations (32) and (39);

3. evaluate the observational relations along the past light cone: d_t , ρ_v , N_c , z , V ;
4. fit a straight line with the points obtained by numerical integration, according to equation (40);
5. is the fitting linear and with negative slope?
6. if the answer is no, then choose other functions $f(r)$, $F(r)$, $\beta(r)$ and go all over again; else STOP: a fractal solution was modelled and D and σ are easily found.

The description above is very schematic since many other important details like root finding algorithm for equations (27) and (29), initial conditions for the ODE's, numerical integrator, etc, have to be considered. A full discussion of the numerical problems involved in this method can be found in [41, 42]. In any case, from the description above it is obvious that if the distribution of dust remains homogeneous throughout the past null geodesic, the volume density will not change and $\log \rho_v = \text{constant}$, and $D = 3$. In this case the plot $\log \rho_v$ vs. $\log d_t$ will be a straight line with zero slope.

Before closing this section, a few words are necessary here in order to explain why the luminosity distance d_t was chosen as the measurement of distance in this relativistic approach to fractal cosmology. As is well known, in relativistic cosmology we do not have an unique way of measuring the distance between source and observer since their separation depends on circumstances. We can, for instance, make use of geometrically defined distances like the proper radius, or observationally defined distances like the luminosity distance or the observer area distance (also known as angular diameter distance) in order to say that a certain object lies at a certain distance from us. The circumstances which tell us which definition to use can also be determined on observational grounds, and so if we only have at our disposal the apparent magnitudes of galaxies we associate to each of them the luminosity distance and use such measurement in our analysis. On the other hand, if these apparent magnitudes are corrected by the redshift of the sources, we can then associate the corrected luminosity distance, which is the same as the observer area distance obtained if we have the apparent size of the objects [43], and, therefore, another kind of distance measure is obtained. Any of these observational distances are as valid as any other, as real as any other, with the choice being dictated by the availability of data, the nature of the problem being treated and its convenience, but they will only have the same value at $z \ll 1$, varying sometimes widely for larger z (see [44] for a comparison of these distances in simple cosmological models). In this work we are interested in observables because we seek to compare theory with observations, and this means that geometrical distances are of no interest here. Consequently, the approach of this relativistic cosmological model is different from others where unobservable coordinate distances (differences between coordinates) and separations (integration of the line element dS over some previously defined surface) are taken as measure of distance, and in order to develop a treatment coherent with the observational approach of this problem we need to make a choice among the observational distances based on the nature of the problem and the observations available. It is the intention of this work to propose a relativistic extension of Pietronero-Wertz model in order to describe the observed scaling behaviour of the distribution of galaxies detected by the recent all-sky redshift surveys, and in this area of research it is usual for observers to take the luminosity distance as their indicator of distance (for example, [45] does use d_t in its statistical analysis of a sample of IRAS galaxies). It seems therefore perfectly reasonable to take the luminosity distance as the

most appropriate definition of distance to use in the context of this work, because what is sought is to mimic the current methodology followed by many observers in this field, and to carry out a comparison between the theoretical predictions of this relativistic cosmological model and the observational results brought by the redshift surveys.

5. HOW INHOMOGENEOUS IS A “HOMOGENEOUS” UNIVERSE?

The observational relations presented in the previous section were derived with the clear intention of developing a relativistic fractal model in the sense of Pietronero and Wertz, but as we shall see in this section, the application of these same observational relations to the spatially homogeneous Friedmann spacetime bring to light some very interesting and unexpected results.

In order to see how those results are achieved, let me say first of all that the question which motivated the application of the observational relations above to the Friedmann spacetime came from the realization that fractal dimensions, in the sense of Pietronero and Wertz, were defined in Euclidean spaces, and it is not clear beforehand whether equation (36) will give the value $D = 3$ for the fractal dimension in the Friedmann metric when we are dealing with observables. One can see more clearly this point if we remember that this metric is *spatially* homogeneous, that is, it has constant local densities at constant time coordinates, and when we integrate along the past light cone, going through hypersurfaces of t constant with each one having different values for the proper density, one may argue that D could depart from the value 3 even in a spatially homogeneous spacetime [41]. With this point in mind, we may go even further and ask whether or not even the Friedmann metric could be compatible with a fractal description of cosmology [46], even in the strongly simplified relativistic version of Pietronero-Wertz model presented here. In order to try to answer those questions, it is convenient to start with the analytically feasible Einstein-de Sitter model.

It is shown in the appendix that the Einstein-de Sitter model can be obtained from Tolman’s spacetime as a special case when

$$f(r) = 1, \quad F(r) = \frac{8}{9}r^3, \quad \beta(r) = \beta_0, \quad (43)$$

where β_0 is a constant. In this case the two differential equations (32) and (39) are easily integrated from $I = t = r = 0$ to $I(r)$ and $t(r)$, respectively yielding

$$1 - I = \left(\frac{3\beta_0^{1/3} - r}{3\beta_0^{1/3}} \right)^2, \quad (44)$$

$$3(t + \beta_0)^{1/3} = 3\beta_0^{1/3} - r. \quad (45)$$

Therefore, along the past light cone the solution (25) of the field equation and its derivatives take the form

$$R = \frac{r}{9}(3\beta_0^{1/3} - r)^2; \quad R' = \frac{1}{9}(3\beta_0^{1/3} - r)^2; \quad \dot{R}' = 2(3\beta_0^{1/3} - r)^{-1}. \quad (46)$$

With the results above, we can easily obtain the observational relations in the Einstein-de Sitter model. The redshift, luminosity distance, cumulative number count, observed volume, volume and local densities are respectively given by

$$1 + z = \left(\frac{3\beta_0^{1/3}}{3\beta_0^{1/3} - r} \right)^2, \quad (47)$$

$$d_t = \frac{9r\beta_0^{4/3}}{(3\beta_0^{1/3} - r)^2}, \quad (48)$$

$$N_c = \frac{2r^3}{9M_G}, \quad (49)$$

$$V = \frac{12\pi r^3 (3\beta_0)^4}{(3\beta_0^{1/3} - r)^6}, \quad (50)$$

$$\rho_v = \frac{(3\beta_0^{1/3} - r)^6}{54\pi(3\beta_0)^4}, \quad (51)$$

$$\rho = \frac{1}{6\pi} \left(\frac{3}{3\beta_0^{1/3} - r} \right)^6. \quad (52)$$

We can see that the local and volume densities change along the past null geodesic, and this is obvious from the dependence on r in both equations (51) and (52). That happens because along the past light cone the integration goes through different surfaces of t constant, and in this sense *the Einstein-de Sitter model does appear to be inhomogeneous along the backward null cone*. The presence of the coordinate r in equation (52) does not mean a spatially inhomogeneous local density, since r is only a parameter along the past null geodesic, and due to equation (45) each value of r corresponds to a single value of t , that is, each r corresponds to a specific $t =$ constant hypersurface given by equation (45). However, as along the null geodesic the local density effectively changes as it goes through different surfaces of constant time coordinate, hence in this sense the model can be thought of as inhomogeneous. The volume density may also change inasmuch as it is being measured through these same hypersurfaces of t constant, where each one has different values for the proper density, and being a cumulative density, ρ_v averages at bigger and bigger volumes in a way that adds more and more different local densities of each spatial section of the model.

Another interesting result follows if we look at the asymptotic limit of the equations above. As the function $\beta(r)$ determines the local time at which $R = 0$ (see the appendix), the surface $t + \beta = 0$ is a surface of singularity, which means that the physical region considered is given by the condition $t + \beta > 0$. Considering equation (45), it is obvious that $r = 3\beta_0^{1/3}$ corresponds to the surface of singularity, meaning that the physical region of the model is given by the condition $0 \leq r \leq 3\beta_0^{1/3}$. Therefore, when $r \rightarrow 3\beta_0^{1/3}$ the observational relations breakdown since $z \rightarrow \infty$, $V \rightarrow \infty$, $d_t \rightarrow \infty$, $\rho \rightarrow \infty$ and $\rho_v \rightarrow 0$.

So we can see that the volume density vanishes asymptotically when observables are plotted, and this result is a consequence of the definition adopted here for the volume density since at the big bang singularity hypersurface the observed volume is infinity, but the total mass is finite. Thus we may say that the following limit holds in the Einstein-de Sitter cosmology:

$$\lim_{d_t \rightarrow \infty} \rho_v = 0. \quad (53)$$

Comparing with equation (9), this result means that this relativistic cosmology obeys the zero global density postulate for a pure hierarchy, as defined by Wertz (and conjectured later by Pietronero), and therefore, we may say that *under the appropriate definitions the Einstein-de Sitter model seems to meet all postulates of a pure hierarchical model, in the sense of Wertz*.

The zero global density postulate was formulated by Wertz since it is a logical result if one takes the Newtonian version of the de Vaucouleurs density power law to its logical asymptotic limit. However, this result has been repeatedly used as a fundamental reason why the universe cannot be hierarchical (or fractal) at larger scales because this postulate not only supposedly contradicts the spatially homogeneous Friedmann model [37], thought to be the correct cosmological model, but also is considered conceptually unacceptable. For instance, [27] states that “if the universe is really the Friedmann type on large scale (...) the inhomogeneous structure must cease on large enough scale”. It is then clear from equation (53) that this is may be a rather misleading approach to relativistic fractal cosmology, and even to cosmology in general, since we definitively can have an observationally based interpretation of the Friedmann model where it has no well defined average density, and is inhomogeneous, with asymptotically zero global density along the past light cone. Therefore, it seems to be an inescapable conclusion that having or not having zero global density in the model is just a question of interpretation.⁹

The inhomogeneity of the Einstein-de Sitter model can be graphically seen in figure 3 where the volume density is plotted against the luminosity distance. At close ranges the model is homogeneous, with the average showing no significant departure from a constant value. This means that at very close ranges the volume density is being measured at our constant time hypersurface, but once the scale increases the average starts to change, which means that ρ_v begins to be calculated at regions where the local density differs from the value at our “now” ($t = 0$). In other words, once the volume density starts to change significantly we would have a significant distancing from the initial time hypersurface, going towards earlier epochs of the model. The plot then shows, in a quantitative way, the scale where theoretically the Einstein-de Sitter model no longer is observationally homogeneous, which in the case of the plot of figure 3 is at about $z \approx 0.04$, or $d_L \approx 160$ Mpc. It is easy to show that a 30% decrease in ρ_v from the value at $t = 0$ (now) happens at about $z \approx 0.1$, or $d_L \approx 500$ Mpc, and therefore, this is approximately the maximum range up to where the homogeneity of the Einstein-de Sitter model can be observed. In other words, as this limit is obtained by solving the past null geodesic equation and using the result in observational relations, we have here a clear evidence that relativistic effects become important in cosmology at very close ranges, and this offers us an observationally based methodology that in principle allows us to ascertain quantitatively such ranges in different scenarios.

Another interesting aspect that comes out of the analysis of figure 3 is the fact that the fractal dimension of the distribution has only the value $D = 3$ in the homogeneous, or flat region of the plot. Beyond this it starts to decrease, effectively making the thinning rate in equation (37) a function of position which, in other words, is incompatible with a single fractal description for the distribution of dust in the Einstein-de Sitter cosmology.

The open and recolapsing Friedmann models also behave in a similar inhomogeneous manner as compared to the Einstein-de Sitter model, although the ranges where the departure from homogeneity starts are different. Figure 4 shows the numerical

⁹ It is interesting to note that many researchers are quick to reject any cosmology with a vanishing global density, although, it could be argued, that the no lesser strange result of an infinity local density at the big bang is accepted without argument. On this respect it also could be argued that both results seem to be exactly what they are, that is, at their face values they are just limits, taken under different circumstances, where the observables breakdown.

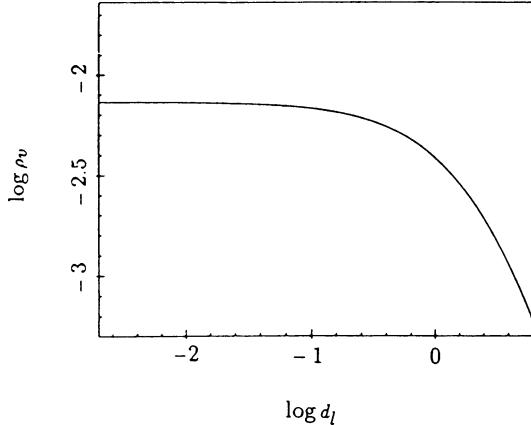


Figure 3. Plot of ρ_v vs. d_ℓ in the Einstein-de Sitter model for the range $0.001 \leq r \leq 1.5$ and with $\beta_0 = 2.7$ (distance is given in Gpc, and $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is assumed). The distribution does not appear to remain homogeneous along the null geodesic and the fractal dimension departs from the initial value 3 [46].

solution for the observational relations in the $K = -1$ Friedmann model, and figure 5 shows the same plot for the recolapsing model [42]. We can clearly see the deviation from homogeneity in both models, although the former starts to deviate at about $d_\ell \approx 250$ Mpc while in the latter that happens at $d_\ell \approx 30$ Mpc. Due to the similarity of the graphs, it seems reasonable to conclude that all Friedmann models appear to obey Wertz's zero global density postulate.

6. TOLMAN FRACTAL SOLUTIONS

Once the observational relations and the numerical methodology for finding fractal solutions in the Tolman model is developed, we can go to the stage of actually specializing the free functions of Tolman's metric in order to see which ones, if any, do have fractal behaviour.¹⁰ Nonetheless, some criteria must be met by those solutions such that some essential observational constraints are obeyed. Those criteria for choosing and accepting the solutions can be listed as follows:

- linearity of the redshift-distance relation for $z < 1$;
- the Hubble constant within the currently accepted range $40 \text{ km s}^{-1} \text{ Mpc}^{-1} < H_0 < 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$;

¹⁰ The numerical code written in FORTRAN 77, and used to solve the past null geodesic in the Tolman metric and to find fractal solutions, is published in [47].

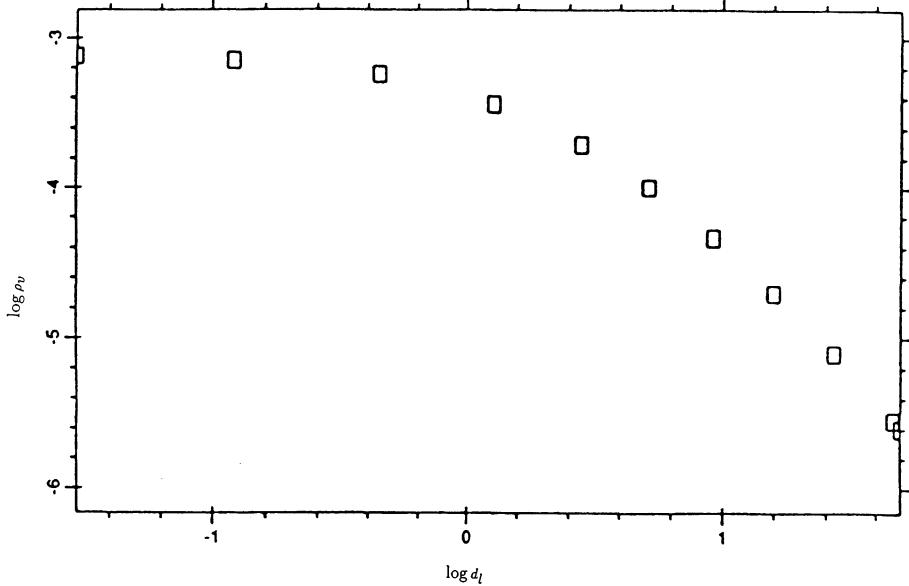


Figure 4. Numerical results for ρ_v vs. d_L in the open Friedmann model model for the range $0.001 \leq r \leq 1.5$ and with $\beta_0 = 3.6$ and $\Omega_0 \approx 0.2$. The deviation from a homogeneous initial region is also visible, and starts at about $d_L \approx 250$ Mpc [42].

- constraint in the fractal dimension of $1 < D < 2$;
- obedience to the Vaucouleurs' density power law $\rho_v \propto d_L^{-\gamma}$.

By means of the methodology described in section 4, a systematic search for fractal solutions was carried out [42], but only hyperbolic type solutions were found to meet all the criteria outlined above. Here I shall only show the simplest hyperbolic fractal solution obtained, since the other classes of fractal solutions did not meet all criteria above. A complete discussion and presentation of the other solutions can be found in [42].

The particular form of the three functions that led to fractal behaviour in hyperbolic models is as follows:

$$\begin{cases} f = \cosh r, \\ F = \alpha r^p, \\ \beta = \beta_0, \end{cases} \quad (54)$$

where α , p , and β_0 are positive constants. The experience with the numerical simulations show that α must be around 10^{-4} to 10^{-5} and p and β_0 can vary from around 0.5 to 4. Figure 6 shows the power law behaviour of ρ_v vs. d_L of the model formed by functions (54), with a fractal dimension of $D = 1.4$. The straight line fitted according to equation (40) is also clearly visible. Figure 7 is the redshift-distance diagram of the same model where we can see the good linear approximation given by functions (54). The slope of the points gives $H_0 \cong 80 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and it is interesting to note that recent measurements made by two different methods suggest a Hubble constant very close to this value [48]. Actually, for $\beta_0 = 3.6$, the value used to get the results shown

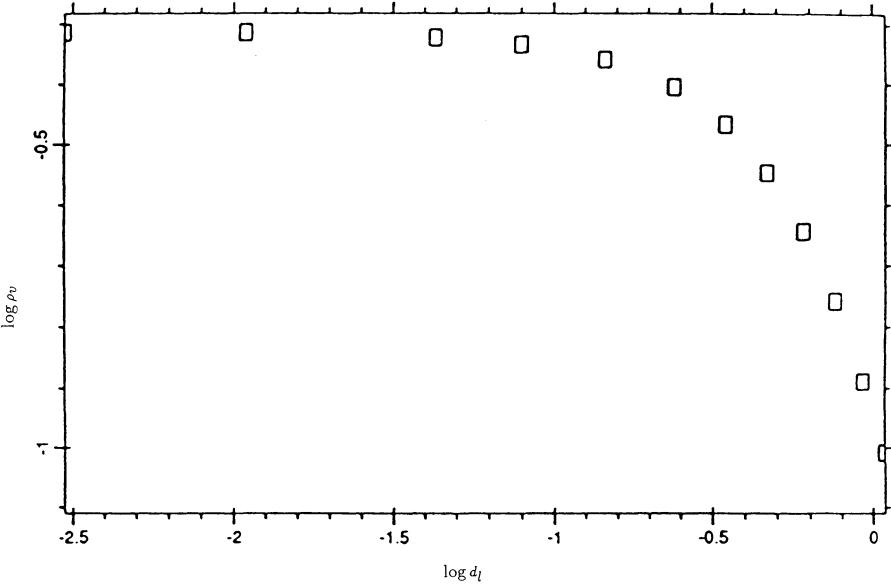


Figure 5. Numerical results for ρ_v vs. d_l in the recollapsing Friedmann model model for the range $0.001 \leq r \leq 1.5$ and with $\beta_0 = 0.7$ and $\Omega_0 \approx 4$. The departure from the homogeneous initial region starts at very close range, at about $d_l \approx 30$ Mpc [42].

in figures 6 and 7, we would have an age of the universe of about 12 Gyr, which is a lower limit if we consider the age of globular clusters [48]. Therefore, also in this point of the age of the universe, the model (54) agrees reasonably well with observations. The integrations with functions (54) were stopped at $z \cong 0.07$, which corresponds to the luminosity distance $d_l \cong 270$ Mpc and this is the redshift depth of the IRAS redshift survey [9]. Finally, figure 8 shows the results for cumulative number counting vs. redshift produced by the model under consideration.

As mentioned above, there are strict limits on the values of the parameters α , p and β_0 , and by means of numerical experimentation it was found that outside the ranges stated above, the fractality of the model (54) is destroyed. Therefore, it seems that this fractal model is *structurally fragile* as a variation of the parameters of the model can produce a qualitative change in its behaviour.

As closing remarks, it can be shown [42] that the model (54) remains fractal at different epochs, with a remarkable constancy in the fractal dimension. In addition, if a Friedmann metric is joined to the solution (54), we can use the internal Tolman solution to find the Friedmann model which best fits another cosmological model that gives a realistic representation of the universe (in this case, the Toman model given by equations [54]), including all inhomogeneities down to some specified length scale mandatory. By doing this procedure, it can be shown [42] that the best Friedmann model is an open one, with $H_0 = 83 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $\Omega_0 \cong 0.002$. This low value for Ω_0 is explained as due to the fact that in this approach no kind of dark matter was considered, but only the luminous matter associated with the galaxies, which are

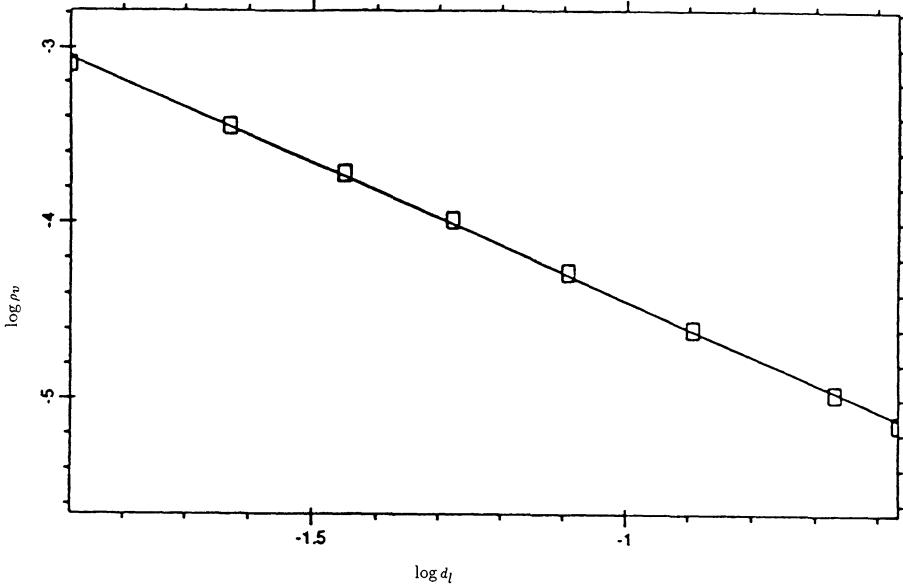


Figure 6. Results of the volume density ρ_v vs. luminosity distance d_l of the hyperbolic model (54). The integration is in the interval $0 \leq r \leq 0.07$ and the constants are $\alpha = 10^{-4}$, $p = 1.4$, $\beta_0 = 3.6$. The fitting coefficients calculated are $a_1 = -6.0$ and $a_2 = -1.6$, giving a fractal dimension $D = 1.4$ and a lower cutoff constant $\sigma = 8.7 \times 10^5$.

assumed to form a fractal system. Galactic luminous matter gives a value for Ω_0 of the same order of magnitude as the one found above.

Joining an external Friedmann metric to the internal inhomogenous solution was initially thought to be a good way of modelling a possible crossover to homogeneity to the fractal system. However, as the Friedmann metric looks inhomogenous at larger scales when measured along the past light cone, this result seems to imply that this external solution is not really mandatory.

7. CONCLUSIONS AND DISCUSSION

This work presented in essence a different approach for modelling the large scale distribution of galaxies, whose main idea is to assume the non-orthodox, but old principle that the empirically observed self-similarity of this distribution is a fundamental fact to be taken into account in any model that attempts a realistic representation of the distribution of galaxies. Under this philosophy, a simple relativistic model was advanced, model which is essentially a translation to a relativistic framework of the Newtonian hierarchical (fractal) cosmology developed by Wertz and Pietronero. This is a simple exploratory model, which although it had to assume some strong simplifications, it is able to obtain some surprising new results, like the inhomogeneity of the Friedmann model, once the observational relations derived for this fractal model are used in this metric. It also shows that the idea of a vanishing global density, postulated for hierar-

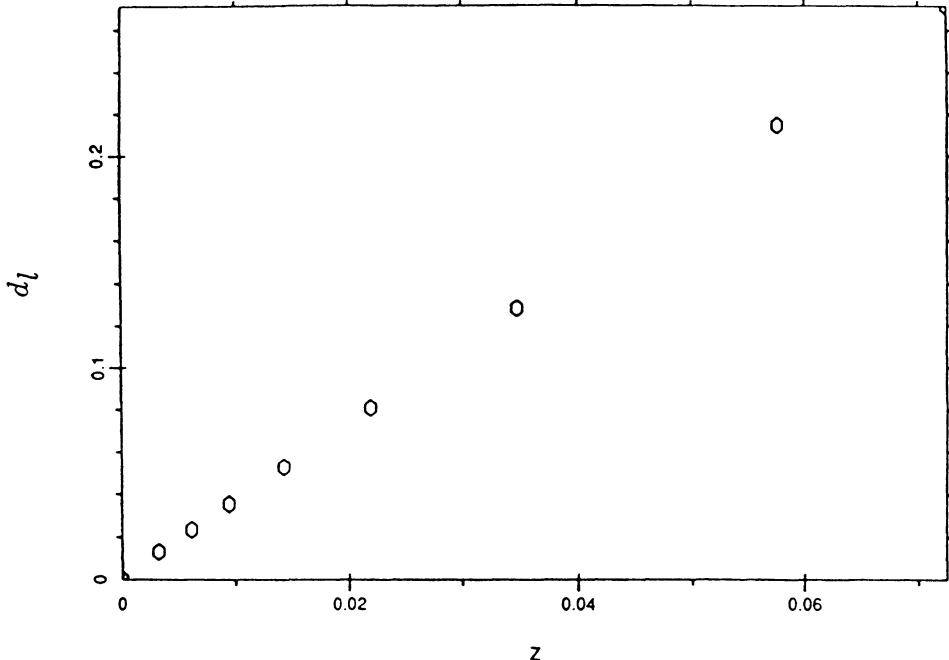


Figure 7. Distance-redshift diagram obtained with the hyperbolic model of functions (54). The slope of the diagram obtained for d_L vs. z gives $H_0 \cong 80 \text{ km s}^{-1} \text{ Mpc}^{-1}$, a value which is within the current uncertainty in the Hubble constant.

chical (fractal) cosmologies, is not necessarily in contradiction with currently accepted cosmological models like the Einstein-de Sitter one, being essentially a question of using the appropriate definitions for density and distance. A general methodology for searching fractal solutions is also advanced, and this methodology is able to successfully find them, showing that solutions of Einstein's field equations approximating a single fractal structure do exist.

In addition to those points, some remarks about the method and the solution should be made. In the first place, the best numerical simulation presented an open, ever expanding model for the large scale distribution of galaxies, and although such class of models are obviously favoured by the simulations, it must be said that flat or recollapsing models are not at all ruled out as there may be some more complex forms for the functions $f(r)$, $F(r)$ and $\beta(r)$ which could produce observationally compatible models, but which were not investigated. Secondly, in view of the results obtained with the fractal inspired observational relations, it seems that observable average densities appear to be physically interesting for the characterization of cosmological models. In particular they are important in the description of a fractal system, and in the determination of the limits of validity of the homogeneous hypothesis. Thirdly, as discussed in the Introduction, the model showed here provides a *description* for the possible fractal structure of the universe, but as in the problem of coastlines, it does not provide an answer to the question of where this fractal system came from, its origins, and why the large scale luminous matter appears to follow a fractal pattern.

To try to answer this last point, it may be crucial a deeper study of the dynamics of

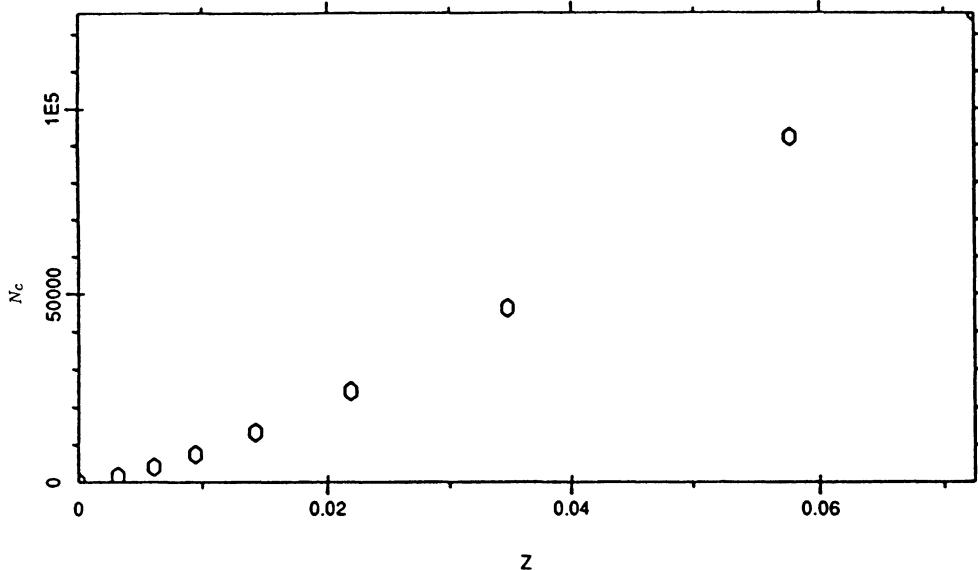


Figure 8. Plot of the results for the cumulative number counting N_c vs. the redshift z given by the integration of the model (54) with the same parameters as in figure 6.

the field equations, or even the particular model showed here, as their nonlinearity may provide some important clues such that one may be able to get closer to answering the question of where this fractal pattern came from. In this respect there are some aspects that may possibly help in this direction. In the first place, structural fragility seems to be a reality in the fractal model discussed (and the others presented in [42]), since there is evidence for this behaviour from the numerical experimentations themselves. Secondly, from the theory of dynamical systems we know that strange attractors have a fractal pattern in phase space. Therefore, the obvious questions is when we see, along the past null geodesic, a self-similar fractal pattern, could that mean that a strange attractor is lurking from behind the scenes? In this case, how can we characterize it? In addition to this, as in some chaotic dynamical systems, could there exists a possible link between the fractal dimension and the Lyapunov exponents in this cosmological context?

Those questions are already in the realm of the second stage of the physicist's approach to dynamical systems, as discussed in the Introduction, and by their own nature, answering them will probably be a very challenging task. In any case, a complete or partial response to any of these points is likely to shed some light on the underlying dynamics of the relativistic field equations in this specific model or, perhaps, in more general ones.

ACKNOWLEDGEMENTS

Thanks go to the organizing committee for this interesting and informative work-

shop, and especially to David Hobill for the smooth way the local organization was carried out.

REFERENCES

- [1] Mandelbrot, B. B., 1983 *The Fractal Geometry of Nature*, (New York: Freeman).
- [2] Pietronero, L., 1987, *Physica A*, **144**, 257.
- [3] Wertz, J. R., 1970, *Newtonian Hierarchical Cosmology*, (PhD thesis), University of Texas at Austin.
- [4] Ellis, G. F. R. and Stoeger, W., 1987, *Class. Quantum Grav.*, **4**, 1697.
- [5] de Lapparent, V., Geller, M. J. and Huchra, J. P., 1986, *Astrophys. J. Lett.*, **302**, L1.
- [6] Kopylov, A. I. *et al* , 1988, *Large Scale Structures of the Universe, 130th IAU Symp.*, ed J Audouze *et al* , (Dordrecht: Kluwer), p 129.
- [7] Geller, M., 1989, *Astronomy, Cosmology and Fundamental Physics*, ed M Caffo *et al* , (Dordrecht: Kluwer), p 83.
- [8] Geller, M. J. and Huchra, J. P., 1989, *Science*, **246**, 897.
- [9] Saunders, W. *et al* , 1991, *Nature*, **349**, 32.
- [10] Ramella, M., Geller, M. J. and Huchra, J. P., 1992, *Astrophys. J.*, **384**, 396.
- [11] Charlier, C. V. L., 1908, *Ark. Mat. Astron. Fys.*, **4**, 1.
- [12] Charlier, C. V. L., 1922, *Ark. Mat. Astron. Fys.*, **16**, 1.
- [13] Feder, J., 1988, *Fractals*, (New York: Plenum).
- [14] Fournier D'Albe, E. E., 1907, *Two New Worlds: I The Infra World; II The Supra World*, (London: Longmans Green).
- [15] Selety, F., 1922, *Ann. Phys.*, **68**, 281.
- [16] Einstein, A., 1922, *Ann. Phys.*, **69**, 436.
- [17] Selety, F., 1923, *Ann. Phys.*, **72**, 58.
- [18] Selety, F., 1924, *Ann. Phys.*, **73**, 290.
- [19] de Vaucouleurs, G., 1970, *Science*, **167**, 1203.
- [20] de Vaucouleurs, G., 1970, *Science*, **168**, 917.
- [21] Wertz, J. R., 1971, *Astrophys. J.*, **164**, 277.
- [22] Haggerty, M. J., 1971, *Astrophys. J.*, **166**, 257.
- [23] Haggerty, M. J. and Wertz, J. R., 1972, *M. N. R. A. S.*, **155**, 495.
- [24] de Vaucouleurs, G. and Wertz, J. R., 1971, *Nature*, **231**, 109.
- [25] Sandage, A., Tamman, G. A. and Hardy, E., 1972, *Astrophys. J.*, **172**, 253.
- [26] de Vaucouleurs, G., 1986, *Gamow Cosmology*, ed F Melchiorri and R Ruffini, (Amsterdam: North-Holland), p 1.
- [27] Feng, L. L., Mo, H. J. and Ruffini, R., 1991, *Astron. Astrophys.*, **243**, 283.
- [28] Bonnor, W. B., 1972, *M. N. R. A. S.*, **159**, 261.
- [29] Wesson, P. S., 1978, *Astrophys. Space Sci.*, **54**, 489.
- [30] Wesson, P. S., 1979, *Astrophys. J.*, **228**, 647.

- [31] Coleman, P. H., Pietronero, L. and Sanders, R. H., 1988, *Astron. Astrophys.*, **200**, L32.
- [32] Davis, M. et al , 1988, *Astrophys. J. Lett.*, **333**, L9.
- [33] Peebles, P. J. E., 1989, *Physica D*, **38**, 273.
- [34] Calzetti, D. and Giavalisco, M., 1991, *Applying Fractals in Astronomy*, ed A Heck and J M Perdang, (Berlin: Springer-Verlag), p 119.
- [35] Coleman, P. H. and Pietronero, L., 1992, *Phys. Reports*, **213**, 311.
- [36] Maurogordato, S., Schaeffer, R. and da Costa, L. N., 1992, *Astrophys. J.*, **390**, 17.
- [37] Ruffini, R., Song, D. J. and Taraglio, S., 1988, *Astron. Astrophys.*, **190**, 1.
- [38] Calzetti, D., Giavalisco, M. and Ruffini, R., 1988, *Astron. Astrophys.*, **198**, 1.
- [39] Pietronero, L., 1988, *Order and Chaos in Nonlinear Physical Systems*, ed S Lundqvist et al , (New York: Plenum Press), p 277.
- [40] Tolman, R. C., 1934, *Proc. Nat. Acad. Sci. (Wash.)*, **20**, 169.
- [41] Ribeiro, M. B., 1992, *Astrophys. J.*, **388**, 1.
- [42] Ribeiro, M. B., 1993, *Astrophys. J.*, **415**, 469.
- [43] Ellis, G. F. R., 1971, *Relativistic Cosmology, Proc. of the Int. School of Physics "Enrico Fermi", General Relativity and Cosmology*, ed R K Sachs, (New York: Academic Press), p 104.
- [44] McVittie, G. C., 1974, *Q. Journal R. A. S.*, **15**, 246.
- [45] Saunders, W. et al , 1990, *M. N. R. A. S.*, **242**, 318.
- [46] Ribeiro, M. B., 1992, *Astrophys. J.*, **395**, 29.
- [47] Ribeiro, M. B., 1992, *On Modelling a Relativistic Hierarchical (Fractal) Cosmology by Tolman's Spacetime*, (PhD thesis), Queen Mary & Westfield College, University of London.
- [48] Peacock, J., 1991, *Nature*, **352**, 378.
- [49] Bondi, H., 1947, *M. N. R. A. S.*, **107**, 410.

APPENDIX. The Friedmann metric as a special case of the Tolman solution

The aim of this appendix is to show how the Tolman solution can be reduced to the Friedmann metric by means of the specializations of the functions $f(r)$, $F(r)$, $\beta(r)$, and the physical differences between the two metrics. This may be useful for those not familiar with the Tolman solution.

It can be shown, by calculating the junction conditions between the Tolman and Friedmann metrics [41], that in order to obtain the latter from the former we have to assume that

$$R(r, t) = a(t) g(r), \quad f(r) = g'(r). \quad (55)$$

By substituting equations (55) into the metric (22) we get

$$dS^2 = dt^2 - a^2(t) \left\{ dr^2 + g^2(r) [d\theta^2 + \sin^2 \theta d\phi^2] \right\}, \quad (56)$$

which is a Friedmann metric if

$$g(r) = \begin{cases} \sin r, \\ r, \\ \sinh r. \end{cases} \quad (57)$$

If we now substitute equations (55) in equation (24) and integrate it, that gives

$$\frac{F}{4} = \frac{4\pi}{3} \rho a^3 g^3. \quad (58)$$

It is worth noting that the time derivative of the equation above gives the well-known relation for the matter-dominated era of a Friedmann universe:

$$\frac{d}{dt} (\rho a^3) = 0.$$

Equation (58) is necessary in order to deduce the usual Friedmann equation. This is possible by substituting equations (55) and (58) into equation (23). The result may be written as

$$\dot{a}^2 = \frac{8\pi}{3} \rho a^2 - K \quad (59)$$

where

$$K = \frac{1 - g'^2}{g^2}.$$

It is easy to see that $K = +1, 0, -1$ if $g = \sin r, r, \sinh r$, respectively, and this shows that equation (59) is indeed the usual Friedmann equation. Let us now write equation (59) in the form

$$\frac{\dot{a}^2 g^2}{2} - \frac{m}{ag} = - (1 - g'^2) \quad (60)$$

where

$$m(r) = \frac{4\pi}{3} \rho a^3 g^3. \quad (61)$$

Equation (60) is interpreted as an energy equation [49] and, in consequence, $m(r)$ is the gravitational mass inside the coordinate r . Thus $4m(r) = F(r)$ and this shows the role of the function $F(r)$ in providing the gravitational mass of the system. In addition, equation (60) shows that the function $f(r)$ in Tolman's spacetime gives the total energy of the system.

The function $\beta(r)$ gives the big bang time of the model, and this can be seen as follows. If "now" is defined as $t = 0$ and if $\beta(r) = 0$, then the hypersurface $t = 0$ is singular, that is, $R = 0$ everywhere¹¹. So $\beta(r)$ gives the age of the universe which in Tolman's spacetime may change if different observers are situated at different radial coordinates r . This is a remarkable departure from Friedmann's model that gives the same age of the universe for all observers on a hypersurface of constant t . In other words, in a Friedmann universe the big bang is simultaneous while in a Tolman one it may be non-simultaneous, that is, the big bang may have occurred at different proper times in different locations. As a consequence another essential ingredient in reducing the Tolman metric to Friedmann is $\beta = \text{constant}$, and so the linkage between β and the Hubble constant is of the form

$$\frac{\dot{R}}{R} = \frac{\dot{a}}{a} = H(t), \quad \text{for } \beta = \beta_0, \quad (62)$$

where β_0 is a constant. Considering equation (25) it is straightforward to conclude that

$$\beta_0 = \frac{2}{3H_0}, \quad (63)$$

¹¹ This is also valid for $f^2 > 1$ and $f^2 < 1$ type solutions of equation (23).

where $H_0 = H(0)$. Equation (63) gives the relationship between β_0 and the Hubble constant H_0 in a Einstein-de Sitter universe.

Bearing this discussion in mind it is easy to see that the usual Friedmann universe requires that $g = \sin r$, r , $\sinh r$ and $F = b_1 \sin^3 r$, $\frac{8}{9}r^3$, $b_2 \sinh^3 r$, which are respectively the cases for $K = +1, 0, -1$. The positive constants b_1 and b_2 are scaling factors necessary to make the density parameter Ω equal to any value different from one in the open and closed models.

Let us now extend equation (63) to the $K = \pm 1$ Friedmann cases. Considering the specializations above, that permit us to get the Friedmann metric from the Tolman solution, and assuming $t = 0$ as our “now”, that is, $t = 0$ being the time coordinate label for the present epoch, it is straightforward to show that the present value H_0 for the Hubble constant in the closed Friedmann model is given by

$$H_0 = \frac{4 \sin 2\Theta_0}{b_1(1 - \cos 2\Theta_0)^2}, \quad (64)$$

where Θ_0 is the solution of

$$4\beta_0 = b_1(2\Theta_0 - \sin 2\Theta_0), \quad (65)$$

and $\beta(r) = \beta_0$ is a constant that gives the age of the universe. In the open Friedmann case we will have then

$$H_0 = \frac{4 \sinh 2\Theta_0}{b_2(\cosh 2\Theta_0 - 1)^2}, \quad (66)$$

and

$$4\beta_0 = b_2(\sinh 2\Theta_0 - 2\Theta_0). \quad (67)$$

For the sake of completeness, let us now obtain the value of the cosmological density parameter Ω in the Friedmann model at the present constant time hypersurface. By definition $\Omega = \rho / \rho_c$ and $\rho_c = 1 / (6 \pi \beta_0^2)$ at $t = 0$. Therefore, in the closed Friedmann model we have that

$$\Omega_0 = \frac{72\beta_0^2}{(b_1)^2(1 - \cos 2\Theta_0)^3}, \quad (68)$$

where Θ_0 is given by equation (65). In the open Friedmann model we have

$$\Omega_0 = \frac{72\beta_0^2}{(b_2)^2(\cosh 2\Theta_0 - 1)^3}, \quad (69)$$

with Θ_0 being the solution of equation (67).

SELF-SIMILAR ASYMPTOTIC SOLUTIONS OF EINSTEIN'S EQUATIONS

A.A. Coley¹ and R.J. van den Hoogen²

Department of Mathematics, Statistics, and Computing Science,
Dalhousie University, Halifax, Nova Scotia, Canada, B3H 3J5

Abstract. The relationship between the existence of self-similar asymptotic solutions of Einstein's equations and equations of state is investigated. For instance, imperfect fluid Bianchi models with 'dimensionless' equations of state are shown to have self-similar asymptotic solutions. Conversely, it is also shown that if the spacetime is self-similar, then the resulting equations of state must be of this same 'dimensionless' form. The conditions under which solutions are asymptotically self-similar are discussed, and it is noted that this is not a generic property of Einstein's equations.

1. INTRODUCTION

How generic are self-similar space-times as asymptotic limits to the Einstein field equations? Wainwright and Hsu [1] have shown for orthogonal Bianchi type A perfect fluid cosmological models that the asymptotic limit points are represented by self-similar models. Similarly, Hewitt and Wainwright [2] have shown for orthogonal Bianchi type B perfect fluid cosmological models that the asymptotic limit points are also represented by self-similar models. Hewitt and Wainwright [3] have also shown that the 'dynamical equilibrium states' are self-similar for the orthogonally transitive G_2 cosmologies. Coley and van den Hoogen [4] have demonstrated that the asymptotic limit points are self-similar for the imperfect fluid Bianchi type V cosmological models with zero cosmological constant when 'dimensionless' equations of state are assumed. We propose to generalize these observations to general orthogonal Bianchi models containing imperfect fluid sources. In section 2, we define self-similarity. In section 3, the Einstein

¹E-mail: aac@cs.dal.ca

²E-mail: hoogen@cs.dal.ca

2. SELF-SIMILARITY

Let (M, g) be a space-time manifold with metric g . A vector \vec{X} that satisfies

$$L_{\mathbf{x}} \mathbf{g} = 2c \mathbf{g}, \quad (1)$$

where c is constant, generates a one parameter family of similarities. If $c = 0$, then \vec{X} is a Killing vector and if $c \neq 0$, then \vec{X} is a homothetic vector. The collection of all similarities of a space-time (M, g) forms a Lie group, called the similarity group. A space-time is defined to be self-similar if it admits a homothetic vector, and transitively self-similar if it admits an H_4 [6]; that is, in addition to the homothetic vector, there exist three Killing vectors that act transitively on 3-dimensional hypersurfaces. In order to be consistent with previous work, we are using the term self-similarity to characterize the properties of the geometry, rather than, as is more conventional, to characterize the properties of the matter [8].

Hsu and Wainwright [6] state a theorem in which the conditions are given for a simply transitive similarity group H_4 to exist.

Theorem. A spacetime (M, g) admits a simply transitive similarity group H_4 if and only if there exists an orthonormal frame $\{\tilde{e}_a\}$ and a scalar field t such that

$$\gamma_{ab}^c = F_{ab}^c t^{-1} \quad (2)$$

$$\vec{e}_a(t) = n_a \quad (3)$$

where F_{ab}^c and n_a are constants.

Proof. See Hsu and Wainwright [6].

3. THE FIELD EQUATIONS

Consider a 4-dimensional space-time manifold (M, g) . Suppose we also have a timelike congruence of curves through every point p of the manifold M . This congruence of curves defines a tangent vector \vec{u} at each space-time point p . The covariant derivative of \vec{u} can be written as

$$u_{a;b} = \frac{\theta}{3} h_{ab} + \sigma_{ab} + \omega_{ab} - \dot{u}_a u_b, \quad (4)$$

where $\sigma_{ab} = \sigma_{(ab)}$, $\sigma_{ab}u^b = 0$, $\sigma_a^a = 0$, $\omega_{ab}u^b = 0$, $\omega_{ab} = \omega_{[ab]}$, and $\dot{u}_a = u_{a;b}u^b$. The above quantities have the following interpretation [7]; θ is the expansion, σ_{ab} is the shear, ω_{ab} is the vorticity, \dot{u}_a is the acceleration, and $h_{ab} = u_a u_b + g_{ab}$ is the projection tensor.

The energy-momentum tensor, T_{ab} , can be decomposed [7] with respect to u_a as follows,

$$T_{ab} = \mu u_a u_b + p h_{ab} + 2q_{(a} u_{b)} + \pi_{ab} \quad (5)$$

where $q_a u^a = 0$, $\pi_{ab}u^b = 0$, $\pi_a^a = 0$ and $\pi_{ab} = \pi_{(ab)}$. In general the quantities μ , q_a , p , and π_{ab} have no physical meaning. However, if u^a is the velocity vector of the fluid, then μ can be interpreted as the energy density, q_a as the energy flux, p as the isotropic pressure, and π_{ab} as the anisotropic stress as measured by an observer moving with 4-velocity u^a .

Assuming that the fluid is moving hypersurface orthogonal, $\vec{u} = \vec{n}$ (\vec{n} is the unit normal to the surfaces of homogeneity), we have that the acceleration and vorticity are zero; that is, $\dot{u}_a = \omega_{ab} = 0$. If we parameterize the surfaces by distance along the geodesics normal to the surfaces, then we can consider the surfaces as surfaces of constant time where $n_a = -t_{,a}$.

Using \vec{u} and the three Killing vectors generating the G_3 group of motions, the Einstein field equations for the orthogonal spatially homogeneous models may be written in terms of an orthonormal tetrad $\{\vec{e}_a\}$. If $\vec{e}_0 = \vec{u}$, then the quantities γ_{ab}^c defined by the commutator relation $[\vec{e}_a, \vec{e}_b] = \gamma_{ab}^c \vec{e}_c$ are spatially independent and are functions of t only. With respect to this basis, the non-zero components of σ_{ab} , q_a , and π_{ab} are respectively $\sigma_{\alpha\beta}$, q_α , and $\pi_{\alpha\beta}$. The quantities γ_{ab}^c may be written in terms of θ , $\sigma_{\alpha\beta}$, and new variables $n_{\alpha\beta}$ and a_β as follows [7]:

$$\begin{aligned} \gamma_{0\alpha}^0 &= \gamma_{\alpha\beta}^0 = 0, \\ \gamma_{0\beta}^\alpha &= -\frac{\theta}{3} h_{\alpha\beta} - \sigma_{\alpha\beta} + \epsilon_{\alpha\beta\tau} \Omega^\tau, \\ \gamma_{\beta\delta}^\alpha &= \epsilon_{\beta\delta\epsilon} n^{\epsilon\alpha} + \delta_\beta^\alpha a_\delta - \delta_\beta^\alpha a_\delta. \end{aligned} \quad (6)$$

Furthermore, the basis $\{\vec{e}_a\}$ can be chosen so that $n_{\alpha\beta} = \text{diag}(n_1, n_2, n_3)$ and $a^\beta = (a, 0, 0)$. The field equations are (equations (113–121), with $\Lambda = 0$ and $\theta_{ab} = \frac{\theta}{3} h_{ab} + \sigma_{ab}$, in MacCallum's Cargèse lectures [7])

$$\dot{\theta} = -\frac{\theta^2}{3} - 2\sigma^2 - \frac{1}{2}(\mu + 3p), \quad (7)$$

$$q_\alpha = 3\sigma_\alpha^\beta a_\beta - \epsilon_{\alpha\beta\gamma} n^\gamma \delta_\beta^\alpha, \quad (8)$$

$$\begin{aligned} \dot{\sigma}_{\alpha\beta} &= -\frac{2}{3} \delta_{\alpha\beta} \left(\frac{\theta^2}{3} - \sigma^2 - \mu \right) + \pi_{\alpha\beta} - \theta \sigma_{\alpha\beta} \\ &\quad - 2\sigma^\gamma_{(\alpha} \epsilon_{\beta)\gamma} \Omega^\delta + 2\epsilon_{\gamma\delta(\alpha} n_{\beta)}^\gamma a^\delta - 2n^\gamma_{(\alpha} n_{\beta)\gamma} \\ &\quad + n_\alpha^\alpha n_{\alpha\beta} + \delta_{\alpha\beta} \left(2a_\gamma a^\gamma + n^\gamma \delta_{\gamma\delta} - \frac{(n_\alpha^\alpha)^2}{2} \right). \end{aligned} \quad (9)$$

The Jacobi identities are

$$\dot{a}_\alpha = -\sigma_\alpha^\beta a_\beta - \frac{\theta}{3} a_\alpha + \epsilon_{\alpha\beta\gamma} a^\beta \Omega^\gamma, \quad (10)$$

$$\dot{n}^{\alpha\beta} = 2n^\gamma (\alpha \epsilon_{\gamma\delta}^\beta) \Omega^\delta + 2n_\gamma^{(\alpha} \sigma^{\beta)\gamma} + 2n_\gamma^{(\alpha} \delta^{\beta)\gamma} \frac{\theta}{3} - n^{\alpha\beta} \theta. \quad (11)$$

The energy-momentum conservation equations are

$$\dot{\mu} = -(\mu + p)\theta - \pi_{\alpha\beta}\sigma^{\alpha\beta} + 2a^\alpha q_\alpha, \quad (12)$$

$$\begin{aligned} \dot{q}_\alpha &= -\epsilon_{\alpha\beta\gamma}q^\beta\Omega^\gamma - \sigma_{\alpha\beta}q^\beta - \frac{4}{3}\theta q_\alpha \\ &\quad + 3a_\beta\pi_\alpha^\beta + \pi_\beta^\gamma\epsilon_{\gamma\alpha\delta}n^{\beta\delta}. \end{aligned} \quad (13)$$

The generalized Friedmann equation is

$$\frac{\theta^2}{3} = \sigma^2 + \mu + \frac{1}{2} \left(6a_\alpha a^\alpha + n^{\alpha\beta}n_{\alpha\beta} - \frac{(n_\alpha^\alpha)^2}{2} \right). \quad (14)$$

The quantity Ω^α is essentially the angular velocity of an observer moving with velocity \vec{e}_0 , of the triad $\{\vec{e}_\alpha\}$ with respect to a set of Fermi propagated axes. For models of Bianchi type A, $\Omega^\alpha = 0$, and for models of Bianchi type B, Ω^α is a linear combination of components of the shear tensor, $\sigma_{\alpha\beta}$ [5].

Equations (8) and (14) are first integrals of the system. Hence, the generalized Friedmann equation (14) can be used to define μ and equation (8) can be used to define q_α in terms of the remaining variables. It is important to note that both μ and q_α are homogeneous functions of degree two in their arguments. The remaining equations (7, 9, 10, 11) constitute a dynamical system. The dynamical system (7, 9, 10, 11) is invariant under the transformation

$$\begin{aligned} \theta &\rightarrow \lambda\theta & a_\alpha &\rightarrow \lambda a_\alpha & p &\rightarrow \lambda^2 p \\ \sigma_{\alpha\beta} &\rightarrow \lambda\sigma_{\alpha\beta} & n_{\alpha\beta} &\rightarrow \lambda n_{\alpha\beta} \\ \pi_{\alpha\beta} &\rightarrow \lambda^2\pi_{\alpha\beta} & t &\rightarrow \lambda^{-1}t \end{aligned} \quad (15)$$

This invariance implies that there exists a symmetry in the dynamical system [9]. With the following change of variables

$$\begin{aligned} \Sigma_{\alpha\beta} &= \frac{\sigma_{\alpha\beta}}{\theta}, & A_\alpha &= \frac{a_\alpha}{\theta}, & N_{\alpha\beta} &= \frac{n_{\alpha\beta}}{\theta}, \\ \Theta &= \ln\theta, & \frac{dt}{d\tau} &= \frac{1}{\theta}, \end{aligned} \quad (16)$$

the new evolution equations for $\Sigma_{\alpha\beta}$, $N_{\alpha\beta}$ and A_α become independent of the variable Θ . That is, Θ decouples from the dynamical system describing the evolution of $\Sigma_{\alpha\beta}$, $N_{\alpha\beta}$ and A_α . The dynamical system can be considered as a reduced dynamical system for $\Sigma_{\alpha\beta}$, $N_{\alpha\beta}$ and A_α together with an evolution equation for Θ .

What equations of state for the pressure p and anisotropic stress $\pi_{\alpha\beta}$ are needed to satisfy the conditions that $p \rightarrow \lambda^2 p$ and $\pi_{\alpha\beta} \rightarrow \lambda^2 \pi_{\alpha\beta}$ in equation (15)? In order for the dynamical system described by equations (7, 9, 10, 11) to have a unique solution, the equations of state must be C^1 functions of their arguments. The equations of state for p and $\pi_{\alpha\beta}$ must also be homogeneous functions of degree two, that is,

$$p(\lambda\theta, \lambda\sigma_{\alpha\beta}, \lambda n_{\alpha\beta}, \lambda a_\alpha) = \lambda^2 p(\theta, \sigma_{\alpha\beta}, n_{\alpha\beta}, a_\alpha), \quad (17)$$

and

$$\pi_{\alpha\beta}(\lambda\theta, \lambda\sigma_{\alpha\beta}, \lambda n_{\alpha\beta}, \lambda a_\alpha) = \lambda^2 \pi_{\alpha\beta}(\theta, \sigma_{\alpha\beta}, n_{\alpha\beta}, a_\alpha). \quad (18)$$

In the new variables (16) the equations of state needed are of the form

$$\begin{aligned} P \equiv \frac{p}{\theta^2} &= \frac{p(\theta, \sigma_{\alpha\beta}, n_{\alpha\beta}, a_\alpha)}{\theta^2} \\ &= P\left(\frac{\theta}{\lambda}, \frac{\sigma_{\alpha\beta}}{\lambda}, \frac{n_{\alpha\beta}}{\lambda}, \frac{a_\alpha}{\lambda}\right) \\ &= P(\Sigma_{\alpha\beta}, N_{\alpha\beta}, A_\alpha) \end{aligned} \quad (19)$$

and

$$\Pi_{\alpha\beta} \equiv \frac{\pi_{\alpha\beta}}{\theta^2} = \Pi_{\alpha\beta}(\Sigma_{\alpha\beta}, N_{\alpha\beta}, A_\alpha). \quad (20)$$

Thus any C^1 functions of the dimensionless variables $(\Sigma_{\alpha\beta}, N_{\alpha\beta}, A_\alpha)$ gives rise to equations of state that allows the reduced dynamical system to remain autonomous in the new variables. We note that $P = p\theta^{-2}$ and $\Pi_{\alpha\beta} = \pi_{\alpha\beta}\theta^{-2}$ are dimensionless and we shall call the corresponding equations of state (19, 20) ‘dimensionless’ equations of state [10].

The singular points of the new reduced dynamical system can be found. At these singular points $\Sigma_{\alpha\beta}$, $N_{\alpha\beta}$, and A_α are constant and consequently the equation

$$\frac{\dot{\theta}}{\theta^2} = -\frac{1}{3} - \Sigma_{\alpha\beta}\Sigma^{\alpha\beta} - \frac{1}{2}(\mu\theta^{-2} + 3P) \quad (21)$$

may be integrated where $P = P(\Sigma_{\alpha\beta}, N_{\alpha\beta}, A_\alpha)$. [Note that equation (14) is used to define $\mu\theta^{-2}$ in terms of the other variables.] The solution to (21) is

$$\theta = \begin{cases} \theta_o t^{-1} & \text{if } -\frac{1}{3} - \Sigma_{\alpha\beta}\Sigma^{\alpha\beta} - \frac{1}{2}(\mu\theta^{-2} + 3P) \neq 0 \\ \theta_o & \text{if } -\frac{1}{3} - \Sigma_{\alpha\beta}\Sigma^{\alpha\beta} - \frac{1}{2}(\mu\theta^{-2} + 3P) = 0 \end{cases} \quad (22)$$

at the singular point. If the condition $\mu + 3p \geq 0$ is satisfied at the singular points then $\theta = \theta_o t^{-1}$ and the remaining physical variables may also be integrated to yield $\sigma_{\alpha\beta} = (\sigma_{\alpha\beta})_o t^{-1}$, $n_{\alpha\beta} = (n_{\alpha\beta})_o t^{-1}$, and $a_\alpha = (a_\alpha)_o t^{-1}$, where the subscript ‘o’ denotes constant values. These solutions imply that the commutation coefficients γ_{ab}^c are inverse functions of t [see equations (6)]. Therefore, using the theorem stated in section 2, the singular points of the reduced system represent transitively self-similar cosmological models. However, the singular points of the reduced dynamical system are the asymptotic limits to the Einstein field equations, therefore the asymptotic limits are self-similar.

Conversely, assume that the asymptotic limit points are self-similar whence the theorem in section 2 implies that the commutation functions γ_{ab}^c are inverse functions of t . Therefore the physical variables θ , $\sigma_{\alpha\beta}$, $n_{\alpha\beta}$, and a_α are also inverse functions of t [see equations (6)]. Equations (7–14) then imply that the pressure p and anisotropic stress $\pi_{\alpha\beta}$ are inverse square functions of t ; that is, $p(t) = p_o t^{-2}$ and $\pi_{\alpha\beta}(t) = (\pi_{\alpha\beta})_o t^{-2}$.

If the pressure p has an equation of state of the form

$$p = p(\theta, \sigma_{\alpha\beta}, n_{\alpha\beta}, a_\alpha), \quad (23)$$

then

$$\begin{aligned} p(t) &= p(\theta(t), \sigma_{\alpha\beta}(t), n_{\alpha\beta}(t), a_\alpha(t)), \\ &= p((\theta)_o t^{-1}, (\sigma_{\alpha\beta})_o t^{-1}, (n_{\alpha\beta})_o t^{-1}, (a_\alpha)_o t^{-1}). \end{aligned} \quad (24)$$

Thus, it follows that

$$\begin{aligned} p(\lambda^{-1}t) &= p(\lambda((\theta)_o t^{-1}), \lambda((\sigma_{\alpha\beta})_o t^{-1}), \lambda((n_{\alpha\beta})_o t^{-1}), \lambda((a_\alpha)_o t^{-1})), \\ &= p(\lambda\theta, \lambda\sigma_{\alpha\beta}, \lambda n_{\alpha\beta}, \lambda a_\alpha). \end{aligned} \quad (25)$$

But $p(\lambda^{-1}t) = \lambda^2 p_o t^{-2} = \lambda^2 p(t)$, thus the equation of state for p is of the form $p(\lambda\theta, \lambda\sigma_{\alpha\beta}, \lambda n_{\alpha\beta}, \lambda a_\alpha) = \lambda^2 p(\theta, \sigma_{\alpha\beta}, n_{\alpha\beta}, a_\alpha)$. The result is similar for $\pi_{\alpha\beta}$. Therefore, assuming that the spacetime is transitively self-similar, the equations of state for the pressure p and anisotropic stress $\pi_{\alpha\beta}$ must be homogeneous functions of degree two in their arguments. The previous two results are summarized in the following theorem.

Theorem. Let there be a G_3 group of isometries acting transitively on a 3-dimensional hypersurface, and assume that the fluid is moving hypersurface orthogonal. Then the asymptotic limits of the Einstein field equations are transitively self-similar if and only if the equations of state for the pressure p and anisotropic stress $\pi_{\alpha\beta}$ are homogeneous functions of degree two; that is,

$$\begin{aligned} p(\lambda\theta, \lambda\sigma_{\alpha\beta}, \lambda n_{\alpha\beta}, \lambda a_\alpha) &= \lambda^2 p(\theta, \sigma_{\alpha\beta}, n_{\alpha\beta}, a_\alpha) \\ \pi_{\alpha\beta}(\lambda\theta, \lambda\sigma_{\alpha\beta}, \lambda n_{\alpha\beta}, \lambda a_\alpha) &= \lambda^2 \pi_{\alpha\beta}(\theta, \sigma_{\alpha\beta}, n_{\alpha\beta}, a_\alpha) \end{aligned} \quad (26)$$

provided $\mu + 3p \geq 0$.

4. SCALAR FIELD

The addition of a scalar field ϕ with potential $V(\phi) = \lambda_n \phi^{2n}$ is equivalent to

$$\mu = \mu_f + \mu_\phi; \quad \mu_\phi \equiv \frac{1}{2} \dot{\phi}^2 + V(\phi), \quad (27)$$

$$p = p_f + p_\phi; \quad p_\phi \equiv \frac{1}{2} \dot{\phi}^2 - V(\phi), \quad (28)$$

$$\pi_{\alpha\beta} = f \pi_{\alpha\beta}, \quad (29)$$

where a subscript f denotes the usual fluid components. With the introduction of the field ϕ , one more equation is needed to complete the system, namely the Klein-Gordon equation

$$\ddot{\phi} = -\theta \dot{\phi} - \frac{\partial V(\phi)}{\partial \phi}. \quad (30)$$

The conditions $p \rightarrow \lambda^2 p$ and $\pi_{\alpha\beta} \rightarrow \lambda^2 \pi_{\alpha\beta}$ necessary for the symmetry to exist in the dynamical system (see section 3) imply that $\phi \rightarrow \lambda \phi$ and $V(\phi) \rightarrow \lambda^2 V(\phi)$. Furthermore, $\dot{\phi} \rightarrow \lambda \dot{\phi}$ also implies that $\phi \rightarrow \phi$, but this is incompatible with $V(\phi) \rightarrow \lambda^2 V(\phi)$, except when $V(\phi) \equiv 0$. Therefore, if $V(\phi) \neq 0$ (e.g., massive scalar field) then there does not exist a symmetry in the dynamical system, but if $V(\phi) \equiv 0$ (i.e., massless scalar field) a symmetry does exist.

The existence of this symmetry in the dynamical system again implies that there exists a transformation of variables. Using the variables in equation (16) and the new variable $\dot{\Phi} = \dot{\phi} \theta^{-1}$, the dynamical system is transformed such that one of the equations decouples. The singular points of the new reduced dynamical system can be found. At these singular points $\Sigma_{\alpha\beta}$, $N_{\alpha\beta}$, A_α and $\dot{\Phi}$ are constant. Thus, at these singular points the equation

$$\frac{\dot{\theta}}{\theta^2} = -\frac{1}{3} - \Sigma_{\alpha\beta} \Sigma^{\alpha\beta} - \dot{\Phi}^2 - \frac{1}{2} (\mu_f \theta^{-2} + 3P_f) \quad (31)$$

may be integrated. [Note that equation (14) can be used to define $\mu_f \theta^{-2}$ in terms of the other variables.] The solution to (31) is

$$\theta = \begin{cases} \theta_o t^{-1} & \text{if } -\frac{1}{3} - \Sigma_{\alpha\beta} \Sigma^{\alpha\beta} - \dot{\Phi}^2 - \frac{1}{2} (\mu_f \theta^{-2} + 3P_f) \neq 0 \\ \theta_o & \text{if } -\frac{1}{3} - \Sigma_{\alpha\beta} \Sigma^{\alpha\beta} - \dot{\Phi}^2 - \frac{1}{2} (\mu_f \theta^{-2} + 3P_f) = 0 \end{cases} \quad (32)$$

at the singular point. If the condition $\mu_f + 3p_f \geq 0$ is satisfied then the remaining physical variables may also be integrated to yield $\sigma_{\alpha\beta} = (\sigma_{\alpha\beta})_o t^{-1}$, $n_{\alpha\beta} = (n_{\alpha\beta})_o t^{-1}$, $a_\alpha = (a_\alpha)_o t^{-1}$ and $\dot{\phi} = (\dot{\phi})_o t^{-1}$. These solutions imply that the commutation coefficients

γ_{ab}^c are inverse functions of t . Therefore, using the theorem stated in section 2 , the singular points represent transitively self-similar cosmological models in the case of a massless scalar field.

Throughout this section it has been implicitly assumed that the singular points occur for finite values of $\dot{\Phi}$. However, there may be singular points as $\dot{\Phi} \rightarrow \pm\infty$. Are such singular points at infinity also self-similar in the case of a massless scalar field? The Klein-Gordon equation (30) may be written in terms of the dimensionless variables and with respect to the new time τ as

$$\frac{d\dot{\Phi}}{d\tau} = -\dot{\Phi} \left(1 + \frac{d(\ln \theta)}{d\tau} \right). \quad (33)$$

This integrates to give $\dot{\Phi} = \kappa \theta^{-1} e^{-\tau}$ (κ is the integration constant). However, as $\dot{\Phi} \rightarrow \pm\infty$ the remaining variables remain finite and

$$\frac{\dot{\theta}}{\theta^2} = \theta^{-1} \frac{d\theta}{d\tau} \sim -\dot{\Phi}^2, \quad (34)$$

which implies that

$$\theta \frac{d\theta}{d\tau} \sim -\kappa^2 e^{-2\tau}. \quad (35)$$

Upon integrating, we find that $\theta \sim e^{-\tau}$ and hence (with respect to coordinate time) $\theta \sim t^{-1}$. Therefore, the solution corresponding to a singular point at infinite values of the variable $\dot{\Phi}$ is self-similar.

5. MODELS WITH NON-SELF-SIMILAR ASYMPTOTIC LIMITS

The self-similarity of the asymptotic limits of the Einstein field equations is not, however, a robust property. For example, self-similarity is broken if any of the following conditions are satisfied:

- A. The equations of state are not of the form (26).
- B. The existence of a scalar field with a non-zero potential $V(\phi)$.
- C. The existence of a cosmological constant Λ .

In addition, self-similarity may be broken if :

- D. The condition $\mu + 3p \geq 0$ is not satisfied.

As an illustration, let us consider a Friedmann-Robertson-Walker (FRW) model with a perfect fluid source and a scalar field ϕ with potential $V(\phi)$. Since the $k = 0$ FRW model occurs as the asymptotic limit of the $k = \pm 1$ FRW models, we shall further simplify our examples by assuming $k = 0$. The solutions examined here are exact $k = 0$ FRW solutions and are not asymptotic solutions.

The field equations are [see equations (27), (28)]

$$\dot{\theta} = -\frac{1}{3}\theta^2 - \frac{1}{2}(\mu_f + \mu_\phi + 3p_f + 3p_\phi), \quad (36)$$

$$\frac{\theta^2}{3} = \mu_f + \mu_\phi, \quad (37)$$

$$\dot{\mu}_f = -\theta(\mu_f + p_f), \quad (38)$$

$$\dot{\mu}_\phi = -\theta(\mu_\phi + p_\phi). \quad (39)$$

Note that equation (39) is equivalent to the Klein-Gordon equation (30).

5.0.1. Equations of State and Non-zero Potential. First, to demonstrate that dimensionless equations of state imply self-similarity consider a dimensionless equation of state of the form

$$\frac{p}{\theta^2} = (\gamma - 1) \frac{\mu}{\theta^2}; \quad \gamma = \text{constant}. \quad (40)$$

Substituting (40) and (37) into (36) we obtain

$$\frac{\dot{\theta}}{\theta^2} = -\frac{\gamma}{2}. \quad (41)$$

The solution is $\theta \sim t^{-1}$, which implies that the model is self-similar.

However, now assume $\mu_f = p_f = 0$ and a non-zero potential $V(\phi) = \frac{1}{2}m^2\phi^2$. This is equivalent to having a non-dimensionless equation of state for p_ϕ , viz.,

$$\frac{p_\phi}{\theta^2} = \frac{\mu_\phi}{\theta^2} - m^2 \frac{\phi^2}{\theta^2}. \quad (42)$$

Substituting (42) into (36) and (37) we obtain

$$\dot{\theta} = -\frac{1}{3}\theta^2 - \dot{\phi}^2 + m^2\phi^2, \quad (43)$$

$$\frac{1}{3}\theta^2 = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}m^2\phi^2. \quad (44)$$

Taking linear combinations of (43) and (44) we obtain

$$\dot{\theta} + \theta^2 = 2m^2\phi^2, \quad (45)$$

$$\dot{\theta} - \frac{1}{3}\theta^2 = -2\dot{\phi}^2. \quad (46)$$

If we assume that the solution is self-similar, then $\theta \sim t^{-1}$ and $\dot{\theta} \sim t^{-2}$. Equations (45) and (46) then imply that both $\phi \sim t^{-1}$ and $\dot{\phi} \sim t^{-1}$. However this is a contradiction and thus the solution cannot be self-similar.

5.0.2. Cosmological Constant. In a similar manner one may include a cosmological constant in the model by setting $\dot{\phi} = 0$, whence $\mu_\phi = V_o$ and $p_\phi = -V_o$. If we consider the vacuum case in which $p_f = \mu_f = 0$, equation (37) may be simplified to give $\theta = \sqrt{3V_o}$. This solution represents the de Sitter model which is not self-similar. Note that in this case $\mu + 3p = -2V_o$ (see below).

5.0.3. $\mu + 3p \geq 0$. One may also consider a simple imperfect fluid model with bulk viscosity by putting $\mu_\phi = p_\phi = 0$ and $p_f = p_t - \zeta\theta$, where p_t is the thermodynamic pressure and ζ is the bulk viscosity coefficient. If we assume the equations of state

$$\frac{\zeta}{\theta} = \zeta_o \frac{\mu_f}{\theta^2} \quad \text{and} \quad \frac{p_t}{\theta^2} = (\gamma - 1) \frac{\mu_f}{\theta^2}, \quad (47)$$

then the equation of state for $p_f\theta^{-2}$ is dimensionless. Substituting (47) and (37) into (36) we find that

$$\dot{\theta} = -\frac{1}{2}(\gamma - \zeta_o)\theta^2. \quad (48)$$

The solution for $\gamma \neq \zeta_o$ is $\theta \sim t^{-1}$, which is self-similar. However, for $\gamma = \zeta_o$ the solution is $\theta = \theta_o$ (a constant), which is not self-similar. In this case

$$\begin{aligned} \mu + 3p &= \mu_f + 3p_t - 3\zeta\theta, \\ &= \mu_f + 3\mu_f(\gamma - 1) - 3\gamma\mu_f, \\ &= -2\mu_f, \end{aligned} \quad (49)$$

whence $\mu + 3p \not\geq 0$ if the energy density, μ_f , is non-negative.

6. CONCLUSION

The Einstein field equations with an imperfect fluid source were investigated using techniques from dynamical systems theory and employing methods from the theory of symmetries of differential equations. We assumed a G_3 group of isometries acting transitively on 3-dimensional spacelike hypersurfaces; that is, the models under consideration are the spatially homogeneous Bianchi models. Also, we assumed that the fluid flow is moving hypersurface orthogonal. The equations of state for the pressure p and anisotropic stress $\pi_{\alpha\beta}$ were assumed to be homogeneous functions of degree two in the variables $(\theta, \sigma_{\alpha\beta}, n_{\alpha\beta}, a_\alpha)$, so that a symmetry in the field equations could be used to define new variables. The field equations in these new variables allowed one equation to decouple. The singular points of the reduced system were shown to represent self-similar cosmological models provided $\mu + 3p \geq 0$. Also, it was shown that if the spacetime is transitively self-similar then the equations of state for the pressure p and anisotropic stress $\pi_{\alpha\beta}$ are homogeneous functions of degree two in the variables $(\theta, \sigma_{\alpha\beta}, n_{\alpha\beta}, a_\alpha)$, whence the resulting equations of state in the new variables (16) are ‘dimensionless’.

The Einstein field equations containing an imperfect fluid and a scalar field were also investigated. It was shown that in the case of a massive scalar field the field equations in general did not admit a symmetry that allowed us to define new dimensionless variables. However, in the case of a massless scalar field the field equations admitted a symmetry that did allow new variables to be defined. The dynamical system in the new variables led to one equation decoupling and the singular points of the reduced dynamical system were shown to be transitively self-similar if the condition $\mu + 3p \geq 0$ is satisfied. The solutions for $\dot{\Phi} \rightarrow \pm\infty$ were also shown to be self-similar.

We note that in the literature the most utilized equation of state for the pressure p is the barotropic γ -law equation of state

$$p = (\gamma - 1)\mu. \quad (50)$$

For the spatially homogeneous models, the energy density defined by the Friedmann equation (14) is a homogeneous function of degree two. Thus, employing the theorem in section 3, self-similar asymptotic solutions are to be expected for $\gamma \geq 2/3$ (this is the condition for $\mu + 3p \geq 0$). Hence, as is most common in the literature, if a γ -law equation of state is assumed in the spatially homogeneous perfect fluid models, then the asymptotic solutions are generally going to be self-similar.

The asymptotic behaviour of a class of scalar-tensor theories of gravity can also be analyzed. If the action is of the form

$$S = \int d^4x \sqrt{-g} \left(f(\phi)R - \frac{1}{2}(\nabla\phi)^2 - V(\phi) \right), \quad (51)$$

where f and V are arbitrary functions of ϕ [11], then the asymptotic states will not, in general, be self-similar. However, if $V(\phi) \equiv 0$ (this case includes the Brans-Dicke theory of gravity), then for (zero-curvature) isotropic and spatially homogeneous models a symmetry exists, and with a change of variables and provided the equations of state are of the form (26), the asymptotic limits can be shown to be self-similar independent of the form of $f(\phi)$.

It should be stated clearly that asymptotic self-similarity is not a generic property of cosmological models. For instance, with the existence of a scalar field with a non-

zero potential, a cosmological constant, or non-'dimensionless' equations of state, the self-similarity may be broken.

In closing, we note that the dynamical system admits a symmetry, and this symmetry allows one to define new variables. However, the choice of variables (16) made here is by no means the only choice. Any one of the original dynamical variables may be chosen so that it may decouple from the rest. Also, this sort of analysis may be extended to other cosmological models, in particular the orthogonal G_2 cosmological models [3] and the tilting (non-orthogonal) G_3 models [12].

ACKNOWLEDGEMENTS

This research was funded by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

- [1] Wainwright, J. and Hsu L., 1989, *Class. Quantum Grav.*, **6**, 1409.
- [2] Hewitt C.G., and Wainwright J., 1993, *Class. Quantum Grav.*, **10**, 99.
- [3] Hewitt C.G. and Wainwright J., 1990, *Class. Quantum Grav.*, **7**, 2295.
- [4] Coley A.A. and van den Hoogen R.J., 1994, *submitted to J. Math. Phys.*
- [5] Ellis G.F.R. and MacCallum M.A.H., 1969, *Commun. Math. Phys.*, **12**, 108.
- [6] Hsu L. and Wainwright J., 1986, *Class. Quantum Grav.*, **3**, 1105.
- [7] MacCallum M.A.H., 1973, *Cargèse Lectures in Physics*, vol 6, ed E. Schatzman (New York: Gordon and Breach), p 61.
- [8] Coley A.A. and Tupper B.O.J., 1990, *Class. Quantum Grav.*, **7**, 1961.
- [9] Bluman G.W. and . Kumei S., 1989, *Symmetries and Differential Equations* (New York: Springer-Verlag).
- [10] Coley A.A., 1990, *J. Math. Phys.*, **31**, 1698.
- [11] Steinhardt P.J., 1991, *Proceedings of the Sixth Marcel Grossman Meeting on General Relativity*, ed H. Sato and T. Nakamura (Singapore: World Scientific), p 269.
- [12] King A.R. and Ellis G.F.R., 1973, *Commun. Math. Phys.*, **31**, 209.

NONLINEARLY INTERACTING GRAVITATIONAL WAVES IN THE GOWDY T^3 COSMOLOGY

B. K. Berger^{†§1}, D. Garfinkle^{†§}, V. Moncrief^{‡§}, and C. M. Swift[†]

[†]Physics Department, Oakland University, Rochester, MI 48309, USA

[‡]Physics Department, Yale University, New Haven, CT, 06511, USA

[§]Institute for Theoretical Physics, University of California,
Santa Barbara, CA, 93106, USA

Abstract. Nonlinear interactions between the two polarizations of the gravitational waves in the Gowdy T^3 cosmology cause the generation of small-scale spatial structure from smooth initial data as the singularity is approached. The growth of this structure competes with the freezing of the spatial profile characteristic of the velocity dominated approach to the singularity conjectured for these models. Some properties of this phenomenon will be discussed.

1. INTRODUCTION

The Gowdy T^3 cosmology [7, 8, 15] has been used as a test case for a larger program of numerical investigation of the approach to the singularity of cosmologies on $T^3 \times R$ with a spatial $U(1)$ symmetry [4]. The singularity theorems of Penrose and Hawking [16, 11, 12] predict that cosmologies have singularities but don't say how these singularities arise or what their structure is. Almost nothing is known about the singularity structure for generic spatially inhomogeneous cosmologies. There are several possibilities:

(1) Asymptotically velocity term dominated (AVTD) [5, 13]: Time derivatives dominate spatial derivatives in Einstein's equations. Each point of the universe evolves independently toward the singularity with fixed local directions of anisotropic collapse.

¹e-mail: berger@vela.acs.oakland.edu

(2) Locally Mixmaster [1, 2, 14]: Here, since the Mixmaster cosmology is not AVTD, one expects local oscillatory behavior with episodes of behavior that reflects the influence of the spatial derivative terms.

(3) Something else that is neither (1) nor (2).

Our goal is to investigate the approach to the singularity numerically for the most general cosmology possible. (Very few analytic results exist and they apply only to very special models.) However, the singularity behavior of generic cosmologies is difficult to study numerically both because a general model is complicated and because, at the singularity, various quantities of interest become singular. To study (rather than avoid) singularities, an extremely accurate and robust numerical algorithm is required. We have found a method ideally suited to the problem that may allow us to approach the singularity closely enough to identify asymptotic behavior. The scheme is designed to capture exactly the AVTD nature of the spacetime singularity if and when this type of behavior is realized and thus to reveal departures from velocity dominance when more general types of singularities such as Mixmaster oscillations arise. Details of the numerical methods can be found in [4].

Application of our methods to the unpolarized Gowdy T^3 cosmology have lent support to the conjecture [9] that the model is AVTD. During the simulations of the approach to the singularity for these models, it has been noticed that initial data which have smooth (e.g. cosine) spatial dependence develop spatial structure on arbitrarily small scales. Nonlinearities in the wave equations cause ever shorter wavelength modes to be generated until the AVTD regime is reached. Once in the AVTD limit, spatial derivatives contribute negligibly to the evolution of the gravitational wave amplitudes so that the spatial profile of these amplitudes no longer changes. Thus there is a competition between the nonlinear growth of structure and the “freezing” of spatial structure in the AVTD limit.

2. THE GOWDY T^3 COSMOLOGY

The Gowdy T^3 cosmology is a plane symmetric universe on $T^3 \times R$ described by a time variable τ and metric variables λ , P , and Q that depend only on the spatial variable $0 \leq \theta \leq 2\pi$ and τ [7, 8, 15]. The metric is

$$ds^2 = e^{\lambda/2} e^{\tau/2} (-e^{-2\tau} d\tau^2 + d\theta^2) + e^{-\tau} [e^P d\sigma^2 + 2e^P Q d\sigma d\delta + (e^P Q^2 + e^{-P}) d\delta^2] \quad (1)$$

where δ and σ are the coordinates of the symmetry plane. The metric components can be interpreted by analogy to linearized gravity [3]. If P and Q are small, the metric components $g_{\mu\nu}$ can be split into a background $\gamma_{\mu\nu}$ and a perturbation $h_{\mu\nu}$. The background metric contains only λ with the form

$$ds_B^2 = -e^{(\lambda-3\tau)/2} d\tau^2 + e^{(\lambda+\tau)/2} d\theta^2 + e^{-\tau} (d\sigma^2 + d\delta^2). \quad (2)$$

At each θ -value the background is an anisotropic universe. The time variable τ measures the area of the σ - δ plane. The perturbation has the form

$$h_{\mu\nu} = e^{-\tau} P \epsilon_{+\mu\nu} + e^{-\tau} Q \epsilon_{\times\mu\nu} \quad (3)$$

which means that P and Q are the amplitudes of the + and \times polarizations of the gravitational waves. Thus the Gowdy T^3 model contains plane symmetric gravitational waves propagating in the θ -direction. In a vacuum, Einstein's equations split into

(1) wave equations for P and Q :

$$\begin{aligned} P_{,\tau\tau} - e^{-2\tau} P_{,\theta\theta} - e^{2P} \left(Q_{,\tau}^2 - e^{-2\tau} Q_{,\theta}^2 \right) &= 0 \\ Q_{,\tau\tau} - e^{-2\tau} Q_{,\theta\theta} + 2 \left(P_{,\tau} Q_{,\tau} - e^{-2\tau} P_{,\theta} Q_{,\theta} \right) &= 0 \end{aligned} \quad (4)$$

and

(2) background equations for λ with P and Q as sources:

$$\begin{aligned} \lambda_{,\theta} - 2(P_{,\theta} P_{,\tau} + e^{2P} Q_{,\theta} Q_{,\tau}) &= 0 \\ \lambda_{,\tau} - [P_{,\tau}^2 + e^{-2\tau} P_{,\theta}^2 + e^{2P} (Q_{,\tau}^2 + e^{-2\tau} Q_{,\theta}^2)] &= 0. \end{aligned} \quad (5)$$

The T^3 spatial topology is implemented by the imposition of periodic boundary conditions.

For convenience, we shall consider only the solution to the wave equations for the gravitational wave amplitudes P and Q . If desired, the wave amplitudes can be used later to obtain the background variable λ by integration. The background equations are respectively the θ -momentum and Hamiltonian constraints. Their decoupling from the wave equations means (a) initial data for P and Q may be freely chosen (as long as the total θ -momentum vanishes) and (b) there is no problem with the preservation of the constraints during the evolution. (Preservation of the constraints is only automatic in the continuum Einstein equations.) Both these issues are problematical for more general spacetimes.

The wave equations for P and Q may be solved exactly in the AVTD limit where the spatial derivatives are neglected. This yields

$$\begin{aligned} P &= -\beta\tau + \ln[\alpha(1 + \zeta^2 e^{2\beta\tau})] \\ Q &= -\frac{\zeta e^{2\beta\tau}}{\alpha(1 + \zeta^2 e^{2\beta\tau})} + \xi \end{aligned} \quad (6)$$

where α , β , γ , and ξ are constant in τ . (We note that $\tau \rightarrow \infty$ is a curvature singularity.) As $\tau \rightarrow \infty$, we find the limit

$$\begin{aligned} P &\rightarrow \beta\tau \\ Q &\rightarrow Q_0 \end{aligned} \quad (7)$$

where Q_0 is a combination of the previous constants. If the Gowdy universe is AVTD, then it should evolve toward this solution with α , β , γ , and ξ different at each value of θ . Thus in the AVTD limit, each spatial point evolves toward the singularity as an independent universe. The AVTD equations are in fact the geodesic equations for the metric [15]

$$dS^2 = dP^2 + e^{2P} dQ^2. \quad (8)$$

Constants of the motion associated with the three Killing vectors of this metric can be found [9]. A combination of these is the geodesic velocity v defined by

$$v = \sqrt{P_{,\tau}^2 + e^{2P} Q_{,\tau}^2} \quad (9)$$

where $v \rightarrow \beta$ as $\tau \rightarrow \infty$.

There are a number of theorems and conjectures related to the asymptotic behavior of the Gowdy model. Isenberg and Moncrief [13] have shown rigorously that

the polarized Gowdy model ($Q \equiv 0$ or its equivalent) is AVTD (with any value of v). Grubišić and Moncrief [9] have conjectured that the unpolarized Gowdy cosmology is AVTD and, for a generic solution (one that is not secretly polarized), $v < 1$ is required. Heuristically, this may be understood since the neglected term $e^{-2\tau+2P}Q_{,\theta}^2$ must remain small when $P \approx v\tau$.

If $Q = 0$, P satisfies the linear equation [3]

$$P_{,\tau\tau} - e^{-2\tau} P_{,\theta\theta} = 0 \quad (10)$$

which has the exact solution

$$P(\tau, \theta) = \sum_{n=0}^{\infty} a_n Z_0(ne^{-\tau}) \cos n\theta \quad (11)$$

where Z_0 is a zero-order Bessel function (and the cosine dependence has been chosen

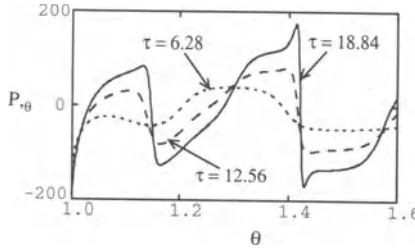


Figure 1. The evolution of structure in $P_{,\theta}$. A feature in $P_{,\theta}$ vs. θ is displayed at three values of the time τ with τ increasing in the direction toward the singularity.

for convenience). If P satisfies the linear wave equation with $Q = 0$, then

$$\bar{P} = \ln(\cosh P), \quad \bar{Q} = \tanh P \quad (12)$$

satisfies the full nonlinear equations. This pseudounpolarized solution has been used for code tests [4].

3. SMALL-SCALE SPATIAL STRUCTURE

To study the conjectured AVTD approach to the singularity, we consider as a generic Gowdy model the initial data

$$P = 0; \quad P_{,\tau} = v_0 \cos \theta; \quad Q = \cos \theta; \quad Q_{,\tau} = 0 \quad (13)$$

where v_0 is the initial value of the previously defined geodesic velocity. Although this choice of initial data seems specific, the evolution appears to be characteristic. For

details see [4]. Since we wish to examine the growth of structure, it is appropriate to examine initial data with the smoothest nontrivial spatial dependence. The wave equations (including the spatial derivatives) are in fact invariant under the transformation $Q \rightarrow \rho Q$, $P \rightarrow P - \ln \rho$ where ρ is a constant. This means that the amplitude of Q is irrelevant as long as it is non-zero. Details of the approach to the singularity for this model (with $v_0 = 10$) are given in [4]. The AVTD behavior arises at various values of θ at different values of τ . Eventually, the model becomes AVTD everywhere.

Generically, the initial $\cos \theta$ dependence quickly generates structures at shorter wavelengths due to the nonlinear coupling of the two polarizations. At large τ , these structures can become narrow and steep (spiky) in a manner reminiscent of shock formation. This is particularly evident in Figure 1 which displays the evolution in τ of $P_{,\theta}$ as the singularity is approached. That this phenomenon is real is demonstrated by running the code with various numbers of spatial grid points. At the highest spatial res-

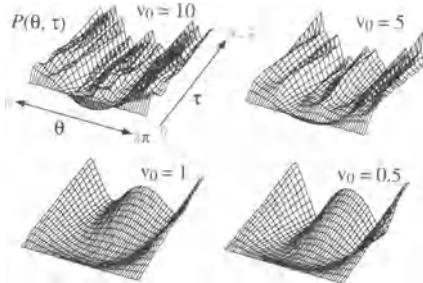


Figure 2. Influence of the parameter v_0 on the growth of small scale spatial structure. The gravitational wave amplitude $P(\theta, \tau)$ is displayed for four simulations which differ only in the initial value of the geodesic velocity v_0 . The initial data are $P = 0$, $P_{,\tau} = v_0 \cos \theta$, $Q = \cos \theta$, and $Q_{,\tau} = 0$.

olution the spiky features are completely resolved. In addition, the solution is correctly followed at a much lower resolution. Where the behavior is not spiky, even the lowest resolution appears to be adequate. Eventually, as the AVTD behavior takes hold, the spatial profile stops changing. Similar small-scale structure is seen with different numerical methods [18], with colliding wavepacket rather than standing wave initial data [6], and in other models containing nonlinear plane waves [17].

Characterization of this structure generation is currently under investigation. Figure 2 shows a comparison of the wave amplitude P for otherwise identical evolutions of our generic initial data which differ only in their value of v_0 . We see that:

- (1) There is a threshold. If the initial $v_0 \leq 1$ everywhere, no small-scale structure evolves.
- (2) The larger v_0 , the greater will be the formation of small-scale structure. This is clearly seen in the contrast between the simulations with $v_0 = 5$ and $v_0 = 10$.

(3) When the solution becomes AVTD, structure formation becomes “frozen.” This is demonstrated clearly in the simulations with smaller values of v_0 .

One therefore expects a v_0 -dependent time scale $\Delta\tau$ to characterize the competition between small-scale structure formation and the approach to the AVTD limit.

The wave equation for P contains competition between the terms $Q_{,\tau}^2$ and $Q_{,\theta}^2$ which cause growth and decay respectively in the wave amplitude. Thus P will grow most where $Q_{,\theta} \approx 0$. This is illustrated in Figure 3 which displays P and Q at two peaks in P . Note that the structure in Q is on a much finer scale (but still well above the computational errors) than that in P . Clearly, peaks in P occur at extrema of Q .

4. CONCLUSIONS

A previous numerical study [4], for reasonably generic Gowdy initial data, has yielded

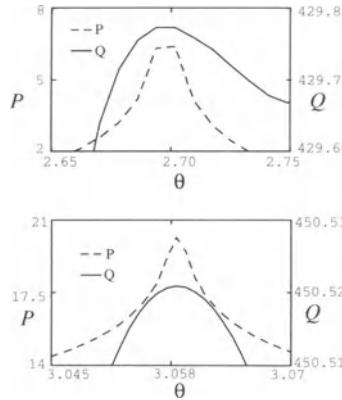


Figure 3. Peaks in P appear at extrema of Q . The upper plot depicts the details of the θ -profile of a peak in P chosen at random compared to $Q(\theta)$ in the same spatial region. Note the scales for Q and θ . (The resolution for this simulation was 800 spatial grid points. The flattening near the peaks is a consequence of limited resolution.) The lower plot depicts the same thing but for the vicinity of a spatial point which has not yet become AVTD late in the evolution. (The resolution for this simulation was 6400 spatial grid points.)

support for the conjecture that these models are AVTD with $v < 1$ everywhere. Grubišić and Moncrief [10] have recently demonstrated detailed agreement with the conjectured [9] form of the decay to the asymptotic regime. An interesting new phenomenon of the development of small-scale spatial structure has been observed. The development of this structure appears to be controlled by the initial value of the geodesic velocity parameter. Further studies to characterize this behavior in terms of the competition between nonlinear generation of short wavelength modes and the freezing of the spatial profile in the AVTD regime are in progress.

ACKNOWLEDGMENTS

This research was supported in part by NSF Grants PHY89-04035 to the Institute for Theoretical Physics, University of California / Santa Barbara, PHY91-07162 and PHY93-0559 to Oakland University, and PHY92-01196 to Yale University. Computations were performed using the facilities of the National Center for Supercomputing Applications at the University of Illinois. BKB wishes to thank the Institute for Geophysics and Planetary Physics at Lawrence Livermore National Laboratory for hospitality.

REFERENCES

- [1] Belinskii, V.A., Khalatnikov, I.M., and Lifshitz, E.M., 1970, *Adv. Phys.*, **19**, 525.
- [2] Belinskii, V.A., Khalatnikov, I.M., and Lifshitz, E.M., 1982, *Adv. Phys.*, **31**, 639.
- [3] Berger, B.K., 1974, *Ann. Phys. (N.Y.)*, **83**, 458.
- [4] Berger, B.K. and Moncrief, V., 1993, *Phys. Rev. D*, in press.
- [5] Eardley, D., Liang, E., and Sachs, R., 1972, *J. Math. Phys.*, **13**, 99.
- [6] Garfinkle, D., 1993, unpublished.
- [7] Gowdy, R.H., 1971, *Phys. Rev. Lett.*, **27**, 826, erratum 1102.
- [8] Gowdy, R.H., 1974, *Ann. Phys. (N.Y.)*, **83**, 203.
- [9] Grubišić, B. and Moncrief, V., 1993a, *Phys. Rev. D*, **47**, 2371.
- [10] Grubišić, B. and Moncrief, V., 1993b, unpublished.
- [11] Hawking, S.W., 1967, *Proc. Roy. Soc. Lond. A*, **300**, 187.
- [12] Hawking, S.W. and Penrose, R., 1970, *Proc. Roy. Soc. Lond. A*, **314**, 529.
- [13] Isenberg, J. and Moncrief, V., 1990, *Ann. Phys. (N.Y.)*, **199**, 84.
- [14] Misner, C.W., 1969, *Phys. Rev. Lett.* **22**, 1071.
- [15] Moncrief, V., 1981, *Ann. Phys. (N.Y.)*, **132**, 87.
- [16] Penrose, R., 1965, *Phys. Rev. Lett.*, **14**, 57.
- [17] Suen, W.M. and Tobias, M., 1993, Workshop on Numerical Relativity, Penn State.
- [18] Swift, C.M., 1993 "Numerical Studies of the Gowdy T^3 Spacetime," Masters Thesis, Oakland University.

BIANCHI IX (MIXMASTER) DYNAMICS

THE MIXMASTER COSMOLOGICAL METRICS

Charles W. Misner ¹

Dept. of Physics, University of Maryland, College Park MD 20742-4111

Abstract. This paper begins with a short presentation of the Bianchi IX or “Mixmaster” cosmological model, and some ways of writing the Einstein equations for it. There is then an interlude describing how I came to a study of this model, and then a report of some mostly unpublished work from a Ph. D. thesis of D. M. (Prakash) Chitre relating approximate solutions to geodesic flows on finite volume negative curvature Riemannian manifolds, for which he could quote results on ergodicity. A final section restates studies of a zero measure set of solutions which in first approximation appear to have only a finite number of Kasner epochs before reaching the singularity. One finds no plausible case for such behavior in better approximations.

1. BIANCHI IX COSMOLOGY

The “Mixmaster” universe is an empty homogeneous cosmology of Bianchi type IX, i.e., it is a (class of) solutions of Einstein’s vacuum equations with three dimensional spacelike surfaces which are orbits of a particular symmetry group. The spacetime manifold M is $R \times S^3 = R^4 - \{0\}$ with the differential structure inherited from the standard euclidean coordinates $wxyz$ of R^4 . When R^4 is presented as quaternions, M is the manifold of all $q \neq 0$ where $q = w + ix + jy + kz$.

The symmetry group is the simply connected covering group of $\text{SO}(3)$ and is realized as the unit quaternions $|p| = 1$ acting as transformations $q \mapsto pq$. These mappings of M onto itself preserve $|q|^2 = \bar{q}q = w^2 + x^2 + y^2 + z^2$ and leave invariant the differential forms

$$\begin{aligned}\sigma_x &= \frac{1}{\bar{q}q}[i\bar{q}dq + (\overline{i\bar{q}dq})] \\ &= \frac{2}{|q|^2}[x dw - w dx - z dy + y dz]\end{aligned}\tag{1}$$

¹e-mail: Misner@umdHEP.umd.edu

and σ_y, σ_z, dt which are defined by similar formulae where i above is replaced by $j, k, 1$ respectively. [Note that quaternion conjugation satisfies $\overline{(pq)} = \bar{q} \bar{p}$ and that $ij = -ji = k$ etc.] From these definitions one can verify $d\sigma_x = \sigma_y \wedge \sigma_z$ and cyclic permutations thereof in which one sees the structure constants of $SO(3)$. One also integrates dt to find that $|q|^2 = e^t$.

The metric on M is taken to be

$$ds^2 = -N^2(t)dt^2 + g_{ij}(t)\sigma_i\sigma_j . \quad (2)$$

It will be invariant under the transformations $q \rightarrow pq$ with $|p| = 1$ since t and the σ_i are. Then the coefficients are parameterized as

$$\begin{aligned} g_{xx} &= \exp[2(-\Omega + \beta_+ + \sqrt{3}\beta_-)] \\ g_{yy} &= \exp[2(-\Omega + \beta_+ - \sqrt{3}\beta_-)] \\ g_{zz} &= \exp[2(-\Omega - 2\beta_+)] \end{aligned} \quad (3)$$

with off-diagonal $g_{ij} = 0$. This lets Ω control the 3-volume $\sqrt{3}g = e^{-3\Omega}$ while the β 's control the shape or shear deformation of the 3-geometry.

1.1. ADM Hamiltonian

The Einstein equations can be summarized in a variational principle $\delta I = 0$ with

$$I = \int(p_+ d\beta_+ + p_- d\beta_- - H d\Omega) \quad (4)$$

where

$$H^2 = p_+^2 + p_-^2 + e^{-4\Omega}(V - 1) \quad (5)$$

and $V = V(\beta_+, \beta_-)$. To reconstruct the metric from solutions of Hamilton's equations one must select a time coordinate. The choice $t = \Omega$ requires that one set $N \propto H^{-1}e^{-3\Omega}$ which can lead to difficulties when $H = 0$ but is not problematic near the singularity $\Omega \rightarrow \infty$. [For careful treatment of the scale factors for coordinates and the action I , see Misner (1969b).] The function $V(\beta)$ which appears here is called the anisotropy potential and is defined by

$$\begin{aligned} V &= \frac{1}{3}e^{-8\beta_+} - \frac{4}{3}e^{-2\beta_+} \cosh 2\sqrt{3}\beta_- \\ &\quad + 1 + \frac{2}{3}e^{4\beta_+}(\cosh 4\sqrt{3}\beta_- - 1) \end{aligned} \quad (6)$$

This function is positive definite and has symmetry under $2\pi/3$ rotations in the β_+ - β_- plane. It can also be written

$$V = \frac{1}{3}\text{tr}(e^{4\beta} - 2e^{-2\beta} + 1) \quad (7)$$

where $g_{ij} = e^{-2\Omega}(e^{2\beta})_{ij}$ for a symmetric traceless matrix β . The most important asymptotic form is

$$V \sim \frac{1}{3}e^{-8\beta_+}, \quad |\beta_-| < -\sqrt{3}\beta_+ \quad (8)$$

which defines the typical potential wall, and

$$V - 1 \sim 16e^{4\beta_+}\beta_-^2 - \frac{4}{3}e^{-2\beta_+}, \quad |\beta_-| \ll 1, \beta_+ > 1 \quad (9)$$

which shows a “corner” opposite the typical potential wall.

1.2. Constrained Hamiltonian

A nicer hamiltonian is obtained by taking Ω as a dynamical variable and introducing τ through $d\tau = d\Omega/H$ as the independent path parameter. Then there are three degrees of freedom and the new hamiltonian is

$$\mathcal{H} = \frac{1}{2}[-p_\Omega^2 + p_+^2 + p_-^2 + e^{-4\Omega}(V - 1)] \quad (10)$$

but the constraint $\mathcal{H} = 0$ must be appended to Hamilton's equations to reproduce the full content of Einstein's equations. The lapse function when τ is used as a coordinate time is then $N = e^{-3\Omega}$. [See (Misner 1972) where the coordinate here called τ was called λ . The present notation agrees with Hobill (1994) from BKL apart possibly from some numerical factor.]

Some simple motions can be read from this form. At fixed β the $e^{-4\Omega}(V - 1)$ term becomes negligible as Ω increases. Neglecting this term in \mathcal{H} makes each p constant with $p_\Omega^2 = p_+^2 + p_-^2$. These are the Kasner epochs where $(d\beta_+/d\Omega)^2 + (d\beta_-/d\Omega)^2 = 1$ since Hamilton's equations give $d\beta_\pm/d\Omega = -p_\pm/p_\Omega$. The potential term in equation (10) is essentially the scalar curvature of the 3-space ${}^3R = -6e^{2\Omega}(V - 1)$ so ignoring it gives the dynamics of the spatially flat Bianchi I solutions. To connect this notation with common parameterizations of the Kasner solutions (Lifshitz and Khalatnikov 1963, Hobill 1994) we must set

$$\begin{aligned} p_1 &= (p_+ + \sqrt{3}p_- + p_\Omega)/3p_\Omega &= -u/(u^2 + u + 1) \\ p_2 &= (p_+ - \sqrt{3}p_- + p_\Omega)/3p_\Omega &= (1 + u)/(u^2 + u + 1) \\ p_3 &= (-2p_+ + p_\Omega)/3p_\Omega &= u(1 + u)/(u^2 + u + 1) \end{aligned} \quad (11)$$

If one chooses evolution toward the singularity so $d\Omega/d\tau > 0$ then when these formulae are valid $p_\Omega = -\sqrt{p_+^2 + p_-^2}$.

From the Hamilton equations $dp_\Omega/d\tau = -\partial\mathcal{H}/\partial\Omega$ and $d\Omega/d\tau = -p_\Omega$ one derives

$$\frac{d}{d\tau} (p_\Omega^2 + e^{-4\Omega}) = 4p_\Omega e^{-4\Omega} V \quad (12)$$

which from $V \geq 0$ shows that approaching the singularity ($p_\Omega < 0$) the positive quantity $(p_\Omega^2 + e^{-4\Omega})$ always decreases but remains constant except when the system point is interacting with a potential wall. Near the singularity ($e^{-4\Omega} \rightarrow 0$) one expects these properties to hold for p_Ω^2 . Rigorous control on the behavior of p_Ω near the singularity would be very useful. In Misner (1970) it is suggested from a quasiperiodic solution that $p_\Omega\Omega$ has only a limited percentage variation. Also easily verified from Hamilton's equation and $\mathcal{H} = 0$ is

$$\frac{d}{d\tau} (p_\Omega e^{2\Omega}) = -2e^{2\Omega}(p_+^2 + p_-^2) \leq 0 \quad (13)$$

which in these variables gives the proof (Rugh 1990, Rugh and Jones 1990) that once the volume of the universe $e^{-3\Omega}$ begins to decrease it can never stop decreasing. For once $-d\Omega/d\tau = p_\Omega$ is negative (volume decrease) the quantity $p_\Omega e^{2\Omega}$ must never increase, so p_Ω can never thereafter become zero.

In a bounce against a single potential wall described by equation (8) there are two constants of motion, p_- and $K \equiv \frac{1}{2}p_+ - p_\Omega$ from which the change in u and in p_Ω can be deduced (Misner 1969b). The location where the potential wall becomes significant is fixed by $p_\Omega^2 e^{4\Omega} = (V - 1)$ which gives

$$\beta_+^\omega = -\frac{1}{2}\Omega - \frac{1}{8} \ln(3p_\Omega^2) \quad . \quad (14)$$

2. EARLY MIXMASTER

My involvement with the Mixmaster cosmologies did not begin with the horizon problem, but with questions of anisotropy in the recent universe. Beginning a sabbatical year in Cambridge, I resumed keeping notebooks following the example of my mentor John Wheeler. One notation (16 Nov '66) indicates the speculations that provoked my attention to anisotropic cosmology.

Last night Faulkner and Strittmatter showed me their blackboard globe ... showing that the $z > 1.5$ quasars (all 13 of them) were all ... near the galactic poles. Although this appears most likely either an ununderstood observational bias, or ... we felt that calculations showing its (probably extreme) implications if interpreted in terms of anisotropic cosmologies, should be carried out, and today set about this.

I initially studied Bianchi Type I examples as later described in (Misner 1968). But my previous acquaintance with Bianchi Type IX in work with Taub (Misner 1963, Misner and Taub 1968) led me to extend these techniques to that case. A notebook entry in the spring (20 Mar '67) begins

This is another curvature computation showing that gravitational fields can, like collisionless radiation, give rise to anisotropy potential energy.

and the next day I recorded a definition of the anisotropy potential of equation (7):

Let us study this anisotropy potential. Define

$$V_g = (1/4) \sum_k (e^{4\beta_k} - 2e^{-2\beta_k} + 1)$$

...

These studies were not published accessibly at that time, but were included in an April 1967 essay that won the third prize in the Babson Gravity Research Foundation Essay competition that spring. Although the analysis included the idea of β colliding with the potential walls, its emphasis was on the decrease of anisotropy during expansion as related to the cosmic microwave radiation and on the refocussing of cosmological theory from measuring the FRW constants to explaining why we live in an FRW universe:

The most accurate observations of this radiation concern its isotropy ... surely deserves a better explanation ...

The traditional problem has been to ... distinguish among the small number of homogeneous and isotropic cosmological solutions

... we propose ... wider problem ... explanation of whatever degree of homogeneity and isotropy the observations ... reveal.

These concerns were repeated in (Misner 1968) where to simplify the presentation the analysis was limited to Bianchi type I. The Mixmaster was published later (Misner 1969a) when I had conjectured that it would contribute to the horizon problem. Several months before this Belinskii and Khalatnikov (1969) had completed work on an independent analysis of these same solutions which appeared in Russian within a month

of my Phys. Rev. Letter. Their work was based on earlier unpublished analyses by E. M. Lifshitz and I. M. Khalatnikov in 1962 on a simpler Bianchi type with only two potential walls which introduced the important description $u \mapsto u - 1$ for change from one Kasner epoch to another and the $u \mapsto 1/u$ change marking the end of this sequence/era. The aim of these Russian works was to elucidate the character or existence of a physical singularity in the generic solution of Einstein's equations.

3. TOWARD GEODESIC FLOW

Let us now return to a Hamiltonian description of the dynamics with the aim of emphasizing the repetitive aspects of the evolution toward the singularity. One can nearly halt the motion of the potential wall by a new metric parameterization:

$$\begin{aligned}\beta_+ &= e^t \sinh \zeta \cos \phi \\ \beta_- &= e^t \sinh \zeta \sin \phi \\ \Omega &= e^t \cosh \zeta\end{aligned}\quad (15)$$

together with a modified independent variable $d\lambda = e^{-2t} d\tau$ which rescales the Hamiltonian so $\mathcal{H}e^{2t} = \tilde{\mathcal{H}}$. The result is a Hamiltonian system based on

$$2\tilde{\mathcal{H}} = -p_t^2 + p_\zeta^2 + (p_\phi / \sinh \zeta)^2 + e^{2t} e^{-4\Omega} (V - 1) . \quad (16)$$

This Hamiltonian is conserved since it is independent of λ and must be initialized as $\tilde{\mathcal{H}} = 0$ to yield Einstein's equations. The potential wall's location is now given by $p_t^2 = e^{2t} e^{-4\Omega} (V - 1)$ which using (8) and (15) becomes

$$\sinh \zeta \cos \phi + \frac{1}{2} \cosh \zeta = \frac{1}{8} e^{-t} [2t - \ln(3p_t^2)] \quad (17)$$

which for large t becomes independent of t . Note that when the $(V - 1)$ term in (16) can be neglected $p_t = \Omega p_\Omega + \beta_+ p_+ + \beta_- p_-$ is constant. But also p_t is constant when this term can be considered independent of t , which occurs for a very steep stationary potential wall, i.e., for large t when a steepness factor e^t appears in the exponents of $\exp(-4\Omega - 8\beta_+) = \exp[-4e^t(\cosh \zeta + 2 \sinh \zeta \cos \phi)]$. In this case the potential term is effectively either zero or infinite crossing a fixed $\zeta(\phi)$ curve and thus without effective t dependence. A rigorous limit on the change possible in p_t would be desirable, but I will work under the expectation that p_t is constant for $t \rightarrow \infty$. Then since Hamilton's equations from (16) give $dt/d\lambda = \partial \tilde{\mathcal{H}} / \partial p_t = -p_t$ we see that asymptotically t is just λ rescaled by $-p_t > 0$.

In his Ph. D. thesis D. M. Chitre (1972a) recognized that the metric $ds^2 = d\zeta^2 + \sinh^2 \zeta d\phi^2$ whose inverse appears in equation (16) was that of the Lobachevsky plane so that results from classical hyperbolic geometry could be employed. In particular, when the potential term in equation (16) can be ignored the solutions will be geodesics on the hyperbolic plane. He employed a useful standard coordinate system on this plane by setting $\sinh \zeta = 2r/(1 - r^2)$ to find $ds^2 = 4(dr^2 + r^2 d\phi^2)/(1 - r^2)^2$ so a complex coordinate $z = x + iy$ could be introduced giving a metric for the (geodesic) motion of the system point

$$ds^2 = \frac{4dz d\bar{z}}{(1 - |z|^2)^2} \quad (18)$$

which is conformally flat. But this metric is invariant under the transformations

$$z \mapsto \frac{az + b}{\bar{a} + \bar{b}z} \quad \text{with} \quad |b| < |a| \quad (19)$$

so these transformations which map the unit disc onto itself are conformal transformations. They also map circles into circles and the unit circle onto itself, with straight lines in the complex plane counted as circles. Using these transformations any geodesic of $d\zeta^2$ can be mapped onto the imaginary axis. Thus the geodesics are the arcs of circles in the complex plane that meet the boundary circle $|z| = 1$ perpendicularly. In these coordinates the location of the potential wall at negative β_+ is given by translating equation (17) to find (with $\phi = \pi$ to locate the wall's intersection x_w with the real axis)

$$x_w = \frac{-2 + \sqrt{3 - f^2}}{1 + f} \quad (20)$$

where

$$f = \frac{1}{4}e^{-t}[2t - \ln(3p_t^2)] \quad . \quad (21)$$

The connection between the parameters (t, x, y) and the metric has by now gotten rather exotic:

$$ds^2 = -e^{4t} \exp(-6e^t \frac{1 + |z|^2}{1 - |z|^2}) d\lambda^2 + g_{ij}(t, z) \sigma_i \sigma_j \quad (22)$$

where $g_{ij}(t, z)$ is defined through equations (3), (15) and $e^{i\phi} \sinh \zeta = 2z/(1 - |z|^2)$. Note that the tz and $t\zeta\phi$ parameterizations of the metric include only a sector $\Omega > \sqrt{\beta_+^2 + \beta_-^2}$ of the complete minisuperspace of kinematically possible 3-geometries. This sector includes all regions known to be dynamically accessible as the singularity $t \rightarrow \infty$ is being approached, but it does not include regions of negative Ω where a maximum of volume expansion can occur. Geodesics $z(\lambda)$ in the Lobachevsky plane can be lifted by $t = -p_t \lambda$ to geodesics in minisuperspace with $d\mathcal{S}^2 = -dt^2 + 4dz d\bar{z} (1 - |z|^2)^{-2}$ where they represent Kasner evolution of the 3-geometry.

But not only the potential-free (Kasner) motion can be described by geodesics in minisuperspace. The bounce from one Kasner epoch to the next can also be handled by these techniques. The potential wall can be idealized as steep and stationary by setting $f = 0$ in equation (20) so p_t is constant even during the bounces. But also, the potential walls are themselves geodesics as drawn in Figure 1. The one described by equation (17) in the stationary approximation is just the circle $|z + 2|^2 = 3$. By the transformation (19) with $a = 1$ and $b = 2 - \sqrt{3}$ Chitre makes this circle become the imaginary axis. The Hamiltonian is now

$$2\tilde{\mathcal{H}} = -p_t^2 + \frac{1}{4}(p_x^2 + p_y^2)(1 - x^2 - y^2)^2 + 2\tilde{\mathcal{V}}(t, x, y) \quad . \quad (23)$$

The approximation of a fixed infinitely steep potential wall makes $\tilde{\mathcal{V}}$ independent of t , and a transformation has set the wall of interest to the y -axis, so $\tilde{\mathcal{V}} = \tilde{\mathcal{V}}(x)$. Thus the bounce is just a specular reflection with p_t and p_y constant while p_x changes sign.

Next Chitre plays the hall of mirrors game. The potential well has become a triangle made up of three equivalent steep walls located on geodesics that are circles in the xy -plane. One wall has been translated into the y -axis. Instead of watching the system point interrupt its geodesic motion by bouncing against this wall, we can equivalently let the geodesic continue but, by an $x \mapsto -x$, $y \mapsto y$ transformation, move the potential well to the other side of the y -axis. Instead of seeing the system point

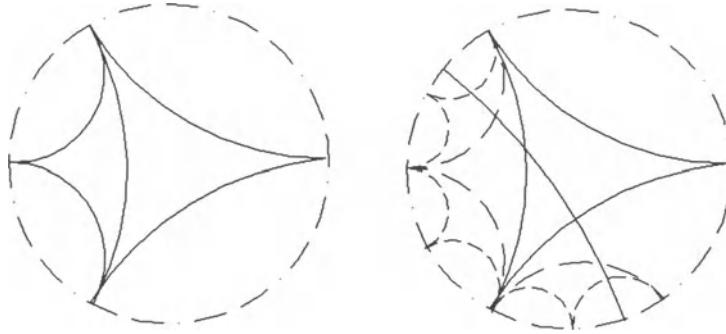


Figure 1. Potential Walls in the Lobachevsky Plane. The shape of a spatial slice of the Mixmaster cosmology is described by the values of a point $z = x + iy$ within the unit disk in the complex plane. The dynamic evolution of this shape is controlled by the Hamiltonian (23). In the sketch of the unit disk above on the left, the right hand geodesic triangle represents the location of the potential walls (at large t , i.e., near the singularity). The smaller geodesic triangle is a mirror image. An isometric transformation (19) taking the common side into a diameter of the unit disk would be needed to make their congruence manifest. In the sketch of the unit disk above on the right further mirror images of these triangles in their sides are shown. Also shown is a geodesic generating the evolution of the universe. When its segments in other images are mapped back into equivalent arcs in the fundamental triangle (potential well), it describes the evolution of the shape of space approaching the singularity.

bounce we watch its mirror image continue the geodesic on the other side of the mirror where it will be headed toward another potential wall in the reflected potential. There it can again continue into another reflection image of the potential well. Since the hyperbolic plane is homogeneous and isotropic, every bounce and every reflection is equivalent under some symmetry to the first one analysed.

The limiting (steep and static) potential well defines a fundamental domain in the Lobachevsky plane which is a triangle bounded by three equivalent geodesics that intersect at infinity ($|z| = 1$). This domain is not compact but has finite area in the $d\zeta^2$ metric. By isometries that give reflections in these walls, one obtains a tiling of the hyperbolic plane and a single geodesic running through an infinity of tiles can be projected back to a succession of bounces in the fundamental domain. Using results from Hopf (1936) and Hedlund (1939) for geodesic flows on quotients of the Lobachevsky plane by a Fuchsian group, Chitre verifies that the flow is ergodic by verifying the hypotheses of theorems showing that the geodesic flow here is metrically transitive and also mixing. To define these terms we need to introduce a measure on the initial conditions or tangent directions for these geodesics. (One assumes $p_t = -1$ in this $t \rightarrow \infty$ limit where the t dependence has by assumption dropped out of the problem.) The tangent directions (x, y, θ) are a point (x, y) in the fundamental domain and a direction θ for the unit tangent vector there. The measure chosen is defined by the integral

$$\int \int \int \frac{4dx dy d\theta}{(1 - x^2 - y^2)^2} \quad (24)$$

and is invariant under the transformations (19) extended to the circle bundle of unit vectors tangent to this manifold. In terms of this invariant measure metrically transitive

means that given any two positive measure measurable sets M and N of tangent directions (initial conditions for the Mixmaster universe near the singularity) the geodesics defined by M after some length λ (evolving toward the singularity) will define tangent directions M_λ of which some coincide with tangent directions in N . Equivalently, metric transitivity means that every measurable set of tangent directions that is invariant under the geodesic flow has either measure zero or full measure (the measure of the whole bundle of tangent directions over the fundamental domain). Mixing means that as a set M of positive measure is evolved along an interval of λ to the set M_λ it will asymptotically ($\lambda \rightarrow \infty$) overlap any measurable set of tangent directions in exactly the ratio of M 's measure that of the whole space (which is 2π times the area of the fundamental domain). Thus after a long interval of $\lambda = t$ any positive measure set of (x, y, θ) tangent directions gets spread out uniformly (although perhaps thinly if the initial set was small) among all such tangent directions. Chitre's summary is

This then characterizes the ‘chaos’ of the universe. If we start with a well-defined state of the universe and let it evolve towards the singularity, we find that it goes through almost all possible anisotropic stages.

4. EVALUATING GEODESIC FLOW

In interpreting the geodesic flows on the Lobachevsky plane, one must be aware that as a plausible approximation to solutions of the Einstein equations only one direction $t \rightarrow +\infty$ on the geodesic can be accepted. The opposite direction toward small t leads to conditions where the motion of the potential wall cannot be neglected. But there are geodesics where even the proper direction leads to problematic situations. Consider the geodesic whose path is the circle $|z - 1 + iR|^2 = R^2$ for $R = \cot(\pi/12) = 2 + \sqrt{3}$ taken as ending at $z = \exp(-5\pi i/6)$ and starting at $z = 1$, points which can be easily identified on the disk on the right in Figure 1. This motion consists of exactly three Kasner epochs delimited by two bounces against potential walls. Likewise, from any point in the fundamental domain there are countably many geodesics that in the forward direction have only a finite number of Kasner epochs. These are the geodesics that end at points on the unit circle where two images of the fundamental potential walls meet. Such points are found by bisecting a finite number of times the arcs on $|z| = 1$ bounded by larger images of the fundamental potential walls. [These trajectories correspond to those in the Belinskii, Khalatnikov, and Lifshitz (1970) description with rational values of u in equations (11).] For these trajectories the hypothesis that only one potential wall at a time need be considered is not convincingly justified since asymptotically the trajectories follow a path between two walls. Although the tangent directions (x, y, θ) giving rise to such trajectories are a set of measure zero (two dimensional in a three dimensional space) it is an intriguing question whether any solution except Taub's has only a finite number of oscillations approaching the singularity.

Let us study in more detail trajectories that have taken their last bounce against a potential wall and are proceeding in the fundamental domain on a geodesic ending at $z = 1$. [These are just the circles $|z - 1 + iR|^2 = R^2$ for any real R with $R^2 > 3$.] To see what has been neglected in this reduction to the Lobachevsky plane let us try to find these trajectories in the description by equation (10). Translating $|z - 1 + iR|^2 = R^2$ back to the $\beta\Omega$ coordinates yields $\Omega - \beta_+ - R\beta_- = 0$ with $R\beta_- > 0$. Then set $p_\Omega = -1$ to normalize the independent variable τ . With the potential term neglected

to give a geodesic we can find p_+ and p_- from $\mathcal{H} = 0$. Two different null geodesics in the minisuperspace metric $d\mathbb{S}^2$ project onto the same geodesic in the Lobachevsky plane, a backward one leading out of the fundamental region with $p_+ < 0$ and the one we want which is $p_+ = 1$, $p_- = 0$, so that β_- remains constant as $\beta_+ \rightarrow +\infty$. These considerations have neglected the potential terms so that we have assumed $2\mathcal{H} = -p_\Omega^2 + p_+^2 + p_-^2$. We will next reinstate some of the potential terms neglected in the geodesic approximation.

As the geodesic of interest retreats from its last collision with a potential wall, the motion is given by $\beta_+ = \tau$ and $\Omega = \tau + R\beta_0$ with $\beta_- = \beta_0$ constant. At $\tau = 0$ the condition $\Omega \gg 1$ assumed for the geodesic approximation then gives $R\beta_0 \gg 1$. In the Hamiltonian (10) the potential term representing the last potential wall encountered is $\frac{1}{6} \exp(-4\Omega - 8\beta_+) = \frac{1}{6} \exp(-12\tau - 4R\beta_0)$ which is readily neglected as τ increases. The potential walls between which the geodesic leads are given by the term $\frac{1}{3} \exp(-4\Omega + 4\beta_+) \cosh(4\sqrt{3}\beta_-) = \frac{1}{3} \exp(-4R\beta_0) \cosh(4\sqrt{3}\beta_0)$ which is independent of τ in this geodesic approximation, but not increasing in significance as would be the case for a geodesic impinging on a potential wall. One can expect then that the neglect of this term is not justified for this trajectory, since even a small force can have a large effect over a long period of time, and the effects of this term on the motion are not projected to ever decrease while the motion is near the geodesic.

To recognize that $-4\Omega + 4\beta_+$ is constant in the geodesic approximation we introduce a new set of parameters $\xi\eta\beta_-$ to replace the previous $\Omega\beta_+\beta_-$. In terms of these new parameters the Hamiltonian is

$$\mathcal{H} = -p_\xi p_\eta + \frac{1}{2}p_-^2 + \mathcal{V} . \quad (25)$$

with

$$\begin{aligned} \mathcal{V} = & \frac{1}{3} \exp(-4\sqrt{2}\eta) [\cosh(4\sqrt{3}\beta_-) - 1] + \frac{1}{6} \exp[-2\sqrt{2}(3\xi - \eta)] \\ & - \frac{2}{3} \exp[-\sqrt{2}(3\xi + \eta)] \cosh(2\sqrt{3}\beta_-) . \end{aligned} \quad (26)$$

The canonical transformation that connects these two forms is

$$\begin{aligned} \xi &= (\Omega + \beta_+)/\sqrt{2}, & p_\xi &= (p_\Omega + p_+)/\sqrt{2}, \\ \eta &= (\Omega - \beta_+)/\sqrt{2}, & p_\eta &= (p_\Omega - p_+)/\sqrt{2} . \end{aligned} \quad (27)$$

In the geodesic approximation upon which we hope to improve these variables have $\sqrt{2}\eta = R\beta_0$ and $\beta_- = \beta_0$ constant while $\sqrt{2}\xi = 2\tau + R\beta_0$. Thus the first term in \mathcal{V} from equation (26) above is constant in this approximation, the second term representing the prior potential wall decreases as $\exp(-12\tau)$ and the third (negative) term which is essential at the maximum of volume expansion decreases as $\exp(-6\tau)$. We will therefore neglect these last two terms and study the motion with

$$\mathcal{V} = \frac{1}{3} \exp(-4\sqrt{2}\eta) [\cosh(4\sqrt{3}\beta_-) - 1] . \quad (28)$$

From the Hamiltonian (25) with the approximate potential (28) one sees that $-p_\xi = d\eta/d\tau$ is constant even without the geodesic ($\mathcal{V} = 0$) approximation. Also neither p_ξ nor $-p_\eta = d\xi/d\tau$ can ever be zero since $\mathcal{H} = 0$ and $\frac{1}{2}p_-^2 + \mathcal{V}$ is positive (apart from the Taub solutions with $p_- = 0 = \beta_-$).

We had chosen $\tau = 0$ to be the stage $\beta_+ = 0$ of the approximate geodesic motion where the previously important term in the potential had decreased to the size of the

term included in equation (28). To continue more accurately from this state we choose as initial conditions at $\tau = 0$

$$\beta_- = \beta_0 , \quad \sqrt{2}\eta = R\beta_0 = \sqrt{2}\xi , \\ p_- = 0 , \quad -p_\eta = \sqrt{2} , \quad -p_\xi\sqrt{2} = \mathcal{V}_0 \equiv \frac{1}{3}\exp(-4R\beta_0)[\cosh(4\sqrt{3}\beta_0) - 1] \quad (29)$$

retaining $-p_\eta = \sqrt{2}$ and $p_- = 0$ from the geodesic motion so that only $-p_\xi$ is changed from zero to a small value which will remain constant during the subsequent evolution. (Recall from the geodesic circles in xy coordinates that $|R| > \sqrt{3}$ while from our choice of $\tau = 0$ at $\beta_+ = 0$ we had $R\beta_0 \gg 1$ so $\mathcal{V}_0 \ll 1$ follows.)

A central question in this study is whether motion under the conditions of validity of the approximation (28) remain valid for $\tau \rightarrow \infty$, or whether motion toward the third potential wall at negative β_+ could evolve. The analysis requires the study of what I call the “nonlinear Bessel equation” (NLB) which in zeroth order is

$$\frac{1}{r} \frac{d}{dr} \left(r \frac{db}{dr} \right) + \sinh b = 0 \quad (30)$$

and has, for small values of b , the solutions $b(r) = Z_0(r)$ where Z_0 is a Bessel function as Chitre (1972b, see also Misner 1970) noticed in studying this question. To see how this equation arises we must study the β_- motion, but first we point out some simpler aspects of this dynamics. In addition to the equation $-p_\xi = \text{const}$ that arises from the absence of ξ from the Hamiltonian (25) in the approximation (28) one finds also that

$$\frac{d}{d\tau}(p_\xi p_\eta) = p_\xi \frac{dp_\eta}{d\tau} = -p_\xi \frac{\partial \mathcal{H}}{\partial \eta} = -4\mathcal{V}_0 \mathcal{V} < 0 \quad (31)$$

where from $\mathcal{H} = 0$ one has

$$p_\eta p_\xi = \frac{1}{2}p_-^2 + \mathcal{V} > 0 . \quad (32)$$

Thus $p_\eta p_\xi = -p_\eta \mathcal{V}_0 / \sqrt{2}$ is a positive quantity which monotonically decreases. The NLB equation appears when the Hamiltonian equations for β_- are written. They read $d\beta_-/d\tau = p_-$ and $dp_-/d\tau = -\partial \mathcal{H}/\partial \beta_-$ or

$$\frac{d^2\beta_-}{d\tau^2} + \frac{4}{\sqrt{3}} \exp(-4\sqrt{2}\eta) \sinh(4\sqrt{3}\beta_-) = 0 . \quad (33)$$

Since $-p_\xi = d\eta/d\tau = \mathcal{V}_0/\sqrt{2}$ we can, using also the initial conditions assumed at $\tau = 0$ write

$$\sqrt{2}\eta = \mathcal{V}_0\tau + R\beta_0 \quad (34)$$

and eliminate η from this differential equation:

$$\frac{d^2\beta_-}{d\tau^2} + \frac{4}{\sqrt{3}} \exp(-4\mathcal{V}_0\tau - 4R\beta_0) \sinh(4\sqrt{3}\beta_-) = 0 . \quad (35)$$

Then a rescaling of both β_- and τ to $b = 4\sqrt{3}\beta_-$ and $\nu = 2\mathcal{V}_0\tau + 2R\beta_0 + \ln(\mathcal{V}_0/2)$ gives

$$\frac{d^2b}{d\nu^2} + \exp(-2\nu) \sinh(b) = 0 . \quad (36)$$

The substitution $r = \exp(-\nu)$ then leads to equation (30) which was first derived in this connection by Belinskii and Khalatnikov (1969). Note that the change from τ to ν

as a natural independent variable corresponds to a change in time scale from $\Delta\tau \approx 1$ for the prior Kasner epochs to $\Delta\nu \approx 1$ or $\Delta\tau \approx 1/2V_0 \gg 1$ for any subsequent oscillations in β_- .

Equation (36) for $\nu \ll -1$ or equation (30) for $r \gg 1$ has a simple behavior of oscillations in an adiabatically changing potential well where the oscillation period is short compared to the time scale $\Delta\nu \approx 1$ on which the potential changes. This condition of negative initial ν arises for many but not all initial conditions, e.g., for $\beta_0^2 \ll 1$ or for $|R| > 2\sqrt{3}$. Under these conditions the amplitude B of the β_- oscillation increases with $B^2 e^{-\nu} \approx \text{const}$ in the (Bessel function) case of small B^2 and with $B^2 \exp(B - 2\nu) \approx \text{const}$ for $B^2 \gg 1$ which is a long era of many Kasner epochs. In all cases one expects that for $\nu \gg 0$ the potential becomes negligible in equation (36), and plausible asymptotic forms have been given by Belinskii and Khalatnikov (1969), but it would be very desirable to have rigorous mathematical results concerning the NLB equation. Solutions where b remains bounded (like the $J_0(e^{-\nu})$ Bessel function solution) can be found as power series in $r = e^{-\nu}$:

$$b = b_0 - \frac{1}{4}r^2 \sinh b_0 + 2^{-7}r^4 \sinh 2b_0 + \frac{1}{3}2^{-10}r^6 \left(\frac{5}{3} \sinh b_0 - \sinh 3b_0 \right) + O(r^8) \quad (37)$$

for which a positive radius of convergence $r^2 < \frac{1}{2} \exp(-|b_0|)$ looks plausible. These solutions become $y = 0$ or $u = 1$ geodesics in the Lobachevsky plane and return again after one bounce against the potential wall at $x = -2 + \sqrt{3}$ to the domain of validity of equation (36) with revised initial conditions. Solutions with $b \propto \nu$ are not known to me in any systematic asymptotic series. An approximation

$$b = 2k(\nu - \nu_o) - \frac{\exp[-2(1 - |k|)\nu - 2|k|\nu_o]}{8(1 - |k|)^2} \quad (38)$$

gives a small value $-\exp[-4(1 - |k|)\nu - 4|k|\nu_o]/16(1 - |k|)^2$ to the left hand side of equation (36) when $\nu \gg 1$ and $|k| < 1$ and thus is a plausible asymptotic form. It corresponds to the start of a new geodesic era. If, as seems likely, every solution of the NLB equation has one of these asymptotic forms for $\nu \rightarrow \infty$ then, apart from the Taub solutions with $b = 0$, every solution returns to geodesic motion in the Lobachevsky plane as the approximation (28) fails when $2\sqrt{2}(\eta - 3\xi) = (1 - k^2)\nu + O(1)$ becomes sufficiently large.

Let us summarize this analysis of the (two dimensional, zero measure) set of tangent directions to the Lobachevsky plane that produce geodesics which reach the singularity after only a finite number of encounters with the potential walls. For these cosmologies the geodesic approximation (or the equivalent Kasner oscillations generated by $u \mapsto u - 1$ and $u \mapsto 1/u$) are not adequate. The $u = \infty$ state does not persist when a better approximation is employed, and in many cases a new regime of Kasner oscillations will begin on a much longer time scale in τ . There is no plausible suggestion that any solution except the Taub solutions can reach the singularity while encountering the potential walls only a finite number of times. To show this more convincingly would require a rigorous qualitative description of all solutions of the NLB equation (30).

As a final comment I note that the measure (24) which allowed ergodic properties to be stated in the Lobachevsky plane approximation arises out of the Einstein equation dynamics. This dynamics is expressed as a Hamiltonian for motion in minisuperspace which is unique up to a conformal factor and displays a similarly unique minisuperspace metric. Approaching the singularity $t \rightarrow \infty$ an asymptotic symmetry of t independence for the potential and conformal metric arises. The minisuperspace metric that displays

this symmetry as a Killing vector is unique and defines the measure we use in stating ergodic properties. If a dynamically defined measure on initial conditions near the singularity existed more generally it would be very significant. For example conjectures in nonquantum relativity such as “for almost all initial conditions at the singularity inflationary evolution leads to regions of spacetime approximating the spatially flat FRW cosmologies” could have a precise meaning. Important questions on quantum geometry such as factor ordering or the choice of an inner product might also become more tractable if a natural measure on initial states were available. Thus further examples would be very useful where the existence of asymptotic symmetries was decidable.

REFERENCES

- Belinskii, V. A., and I. M. Khalatnikov, 1969, On the nature of the singularities in the general solution of the gravitational equations, *Zh. Eksp. Teor. Fiz.* **56**, 1701–1712, *Soviet Physics—JETP* **29**, 911–917.
- Belinskii, V. A., I. M. Khalatnikov and E. M. Lifshitz, 1970, Oscillatory approach to a singular point in the relativistic cosmology, *Adv. Phys.* **19**, 525–573.
- Chitre, D. M., 1972a, *Investigation of Vanishing of a Horizon for Bianchi Type IX (the Mixmaster) Universe* (College Park: University of Maryland Ph. D. thesis).
- Chitre, D. M., 1972b, High-frequency sound waves to eliminate a horizon in the mixmaster universe, *Phys. Rev. D* **6**, 3390–3396.
- Hedlund, G. A., 1936, The dynamics of geodesic flows, *Bull. Am. Math. Soc.* **45**, 241–260.
- Hobill, D., 1994, A brief review of “Deterministic chaos in general relativity”, in this volume.
- Hopf, E., 1936, Fuchsian groups and ergodic theory, *Trans. Am. Math. Soc.* **39**, 299–314.
- Lifshitz, E. M., 1963, and I. M. Khalatnikov, Investigations in relativistic cosmology, *Adv. Phys.* **12**, 185–249.
- Misner, Charles W., 1963, The flatter regions of Newman, Unti, and Tamburino’s generalized Schwarzschild space, *J. Math. Phys.* **4**, 924–937.
- , 1968, The isotropy of the universe, *Astrophys. J.* **151**, 431–457.
- , 1969a, Mixmaster universe, *Phys. Rev. Lett.* **22**, 1071–1074.
- , 1969b, Quantum cosmology. I., *Phys. Rev.* **186**, 1319–1327.
- , 1970, Classical and quantum dynamics of a closed universe, in *Relativity*, Carmeli, Fickler, and Witten, eds., (San Francisco: Plenum Pub. Co.), pp 55–79.
- , 1972, Minisuperspace, in *Magic Without Magic—J. A. Wheeler 60th Anniversary Volume*, J. Klauder, ed., (San Francisco: W. H. Freeman and Co.), pp 441–473.
- Misner, C. W. and A. H. Taub, 1968, A singularity-free empty universe, *Zh. Eksp. Teor. Fiz.* **55**, 233–255, *Soviet Physics—JETP* **28**, 122–133 (1969).
- Rugh, S.E., 1990, *Chaotic Behavior and Oscillating Three-volumes in a Space-Time Metric in General Relativity*, Cand. Scient. Thesis, The Niels Bohr Institute, Copenhagen.
- Rugh, S. E. and B. J. T. Jones, 1990, Chaotic behavior and oscillating three-volumes in Bianchi IX universes, *Phys. Lett. A* **147**, 353–359.

THE BELINSKII-KHALATNIKOV-LIFSHITZ DISCRETE EVOLUTION AS AN APPROXIMATION TO MIXMASTER DYNAMICS

Beverly K. Berger¹

Physics Department, Oakland University, Rochester, MI 48309 USA

Abstract. Mixmaster dynamics is defined to be the evolution of vacuum, diagonal Bianchi IX spatially homogeneous cosmologies. Belinskii, Lifshitz, and Khalatnikov (BKL) developed a discrete approximation to Mixmaster dynamics. Here we analyze their approximation to discuss how well it represents the true solution. We show that sufficiently close to the singularity, the BKL approximation describes all the features of an accurately computed Mixmaster trajectory. Within the BKL approximation, the origination of any sensitivity to initial conditions can be explicitly identified. Since a precise analogy can be drawn between the BKL parameters and variables which describe the true Mixmaster dynamics, it can be argued that the trajectories inherit any chaos that characterizes the approximation. Since the discrete evolution which is the closest analog of the true dynamics is only “marginally chaotic,” recent controversies over whether or not the dynamics is chaotic may be more semantic than substantive.

1. INTRODUCTION

The Mixmaster cosmology [52, 42, 33] provides a concrete example of a solution to Einstein’s equations which exhibits dynamical behavior similar to that observed in well-known chaotic Hamiltonian systems. (See for example [37, 28].) This behavior was first recognized in an extensive series of papers by Belinskii, Khalatnikov, and Lifshitz (BKL) [38, 4, 5, 6, 7, 8, 9]. In their search for the generic behavior of solutions to Einstein’s equations, they focused on Bianchi Types VIII and IX as the most general of the spatially homogeneous cosmologies. Their eventual claim was that the dynamics of

¹e-mail: berger@vela.acs.oakland.edu

these models (dubbed Mixmaster dynamics in the independent work of Misner and co-workers [43, 44, 45, 46, 47, 53, 54]) characterized the local behavior of the most general spatially inhomogeneous cosmology. While this claim is still controversial [3], its pursuit led BKL to perform a thorough analysis of the dynamics of the homogeneous models. It is easiest to discuss the Mixmaster behavior by combining the BKL anisotropic metric scale factors [9] with the minisuperspace (MSS) description due to Misner [43]. Essentially, the approach to the singularity is characterized by a sequence of Kasner epochs (free particles in MSS) related (as Bianchi Type II models [56]) through the scattering from the MSS potential. BKL describe any Kasner solution by a single parameter $1 \leq u \leq \infty$ and show that the sequence of Kasner epochs is given by the map $u_{n+1} = u_n - 1$ for $u_n \geq 2$. When $1 \leq u \leq 2$, the map becomes $u_{n+1} = (u_n - 1)^{-1}$ to regain the original range for u and to define the start of an era. We shall call the map $u_{n+1} = f(u_n)$ for $1 \leq u \leq \infty$ the era-epoch or u -map [41]. The Kasner epoch at the start of the $N + 1$ st era can be related to that at the start of the N th era by the Gauss map $u_{N+1} = (u_N - [u_N])^{-1}$ where $[]$ denotes integer part. BKL recognized that the Gauss map displays sensitivity to initial conditions (SIC) while Barrow [1, 2] later demonstrated that it satisfied all standard criteria for a chaotic map. The era-epoch map is only “marginally chaotic” since a trajectory dependent cutoff is required to define the invariant probability for u [11]. All the SIC in this map occurs during the era change [13].

A controversy arose when several groups independently measured zero for the Liapunov exponents (LE’s) for numerically evolved Mixmaster Einstein equations [26, 31, 16]. This appeared to contradict the definitive chaos of the Gauss map. Now it is understood that traditional measures of chaos cannot be applied uniquely to Mixmaster. In particular, the LE can be evaluated to be either positive or zero depending on the choice of time variable [34, 25]. To make direct connection to the era-epoch map is possible, however, through the introduction of MSS proper time λ which remains constant during any Kasner epoch and changes only (and by the same fixed amount) during every bounce. An analogy can be drawn between the discrete sequence of MSS trajectory angles that characterize successive Kasner epochs at discrete values of λ and the sequence of Kasner parameters at a sequence of integers that number the Kasner epochs [11].

BKL (as improved by Chernoff and Barrow [19]) demonstrated that sufficiently near the singularity the entire content of Mixmaster dynamics can be described (in analogy to the Poincaré map) by a 4-parameter discrete sequence. A variant of this parametrization shows explicitly that all the SIC in both the collapse and expansion directions can be understood in terms of only one of these parameters— u in the direction of collapse and v (satisfying the inverse of the era-epoch map in the collapse direction) in the expansion direction. The rest of the dynamics shows no SIC [13].

This BKL approximate dynamics breaks down near the maximum of expansion where there is no Kasner behavior [10] or if the trajectory is too close to the corner rays of the MSS potential where two potential walls give comparable contributions [9, 48]. Even in the latter case (which becomes less and less likely as the singularity is approached), the BKL approximation is eventually regained in the approach to the singularity [9].

The purpose of this article is then to explore the Mixmaster dynamics within this hybrid BKL-Misner formulation with emphasis on the comparison of the BKL approximation to the true dynamics. First we wish to demonstrate that the features of

Mixmaster dynamics are well-known independent of the words or variables used to describe them. The term “Mixmaster dynamics” shall be used to describe the evolution of generic, vacuum, diagonal Bianchi Type IX cosmologies. We note that the words “evolution” and “cosmology” imply that we are discussing a solution to Einstein’s equations. There exist other classes of solutions with Mixmaster behavior such as Bianchi Type VIII [29] and non-diagonal (rotating) Bianchi Types VIII and IX, including some [53] (but not all [29, 32, 30]) couplings to “matter” which shall be taken to mean non-gravitational fields and sources.

The vacuum, diagonal Bianchi Type IX cosmology can be described either by the metric components (a^2, b^2, c^2) [9] or by the minisuperspace (MSS) variables $(\Omega, \beta_+, \beta_-)$ [43] with

$$\begin{aligned}\alpha &= \ln a = \Omega - 2\beta_+ \\ \zeta &= \ln b = \Omega + \beta_+ + \sqrt{3}\beta_- \\ \gamma &= \ln c = \Omega + \beta_+ - \sqrt{3}\beta_-\end{aligned}\tag{1}$$

Note that $e^{3\Omega}$ is the universe volume, while β_{\pm} describe the anisotropy. MSS is the configuration space of the dynamics with axes β_{\pm}, Ω . In this discussion, we shall consider a number of time variables defined by $dt = NdT$ where t is comoving proper time and N is the lapse associated with time variable T . These are shown in Table 1 (where $2H_I = -p_{\Omega}^2 + p_+^2 + p_-^2$ is the Hamiltonian constraint for Bianchi Type I).

Table 1. Some time variables used to analyze Mixmaster dynamics

T	name	N
t	comoving proper time	1
τ	BKL coordinate time	$e^{3\Omega}$
Ω	volume time	$e^{3\Omega}/ p_{\Omega} $
λ	MSS proper time	$ H_I ^{1/2} e^{-3\Omega}$

Einstein’s equations for all class A Bianchi Types can be found by variation of NH where N is the lapse and H is the Hamiltonian constraint given by

$$2H = -p_{\Omega}^2 + p_+^2 + p_-^2 + \xi \mathcal{U}(\Omega, \beta_{\pm})\tag{2}$$

where p_{Ω}, p_{\pm} are canonically conjugate to Ω, β_{\pm} , \mathcal{U} is the MSS potential (spatial scalar curvature) and ξ is an arbitrary constant. We note that if we use Ω -time and solve $H = 0$ for $H_{ADM} = |p_{\Omega}|$, there remain two degrees of freedom.

The case $\mathcal{U} = 0$ corresponds to the vacuum Bianchi Type I (Kasner) solution. Since p_{Ω}, p_{\pm} are now constants of the motion, the MSS trajectory is a straight line described by a single parameter θ defined by

$$\tan \theta = \frac{\Delta \beta_-}{\Delta \beta_+}\tag{3}$$

with Ω the affine parameter along the trajectory. (Constants to define e.g. the intercept of the line can be absorbed in the definition of the spatial variables.) BKL have described Kasner via [9]

$$\begin{aligned}\alpha &= \Lambda p_1 \tau = \Omega(1 - 2 \cos \theta) \\ \zeta &= \Lambda p_2 \tau = \Omega(1 + \cos \theta + \sqrt{3} \sin \theta) \\ \gamma &= \Lambda p_3 \tau = \Omega(1 + \cos \theta - \sqrt{3} \sin \theta)\end{aligned}\quad (4)$$

where the Kasner indices p_i satisfy $\sum_{i=1}^3 p_i = 1 = \sum_{i=1}^3 p_i^2$. We note that Λ relates τ -time to Ω -time. The Kasner indices are in the ranges

$$-\frac{1}{3} \leq p_1 \leq 0 \leq p_2 \leq \frac{2}{3} \leq p_3 \leq 1 \quad (5)$$

so that as $\tau \rightarrow -\infty$ (the singularity), one scale factor grows while the others contract as $e^{3\Omega} \rightarrow 0$.

BKL introduced a parameter u to describe each Kasner solution. In terms of this parameter, the Kasner indices are

$$p_1 = \frac{-u}{u^2 + u + 1}; \quad p_2 = \frac{u + 1}{u^2 + u + 1}; \quad p_3 = \frac{u(u + 1)}{u^2 + u + 1} \quad (6)$$

with $1 \leq u < \infty$. (Equivalently, one may take $0 < u \leq 1$ with the parametrizations of p_2 and p_3 interchanged.)

To study Mixmaster dynamics, we consider the Bianchi Type IX potential

$$\begin{aligned}\mathcal{U} &= e^{4\Omega} \left[e^{-8\beta_+} + 2e^{4\beta_+} (\cosh 4\sqrt{3}\beta_- - 1) - 4e^{-2\beta_+} \cosh 2\sqrt{3}\beta_- \right] \\ &= e^{4\alpha} + e^{4\zeta} + e^{4\gamma} - 2e^{2(\alpha+\zeta)} - 2e^{2(\zeta+\gamma)} - 2e^{2(\gamma+\alpha)}.\end{aligned}\quad (7)$$

Some contours of \mathcal{U} at fixed Ω are displayed in Fig. 1. As $\Omega \rightarrow -\infty$, $\mathcal{U} \rightarrow 0$ unless an exponent vanishes. The exponents α, ζ, γ do vanish on the potential walls. The walls dominate the remaining potential terms except at the corners where

$$\mathcal{U} \approx e^{4\alpha} + e^{4\zeta} - 2e^{2(\alpha+\zeta)} \quad \text{et cyc.} \quad (8)$$

One of the first 2 terms will dominate unless $\alpha \approx \zeta \approx 0$ when all 3 terms are comparable.

2. PROPERTIES OF A NUMERICAL SOLUTION TO EINSTEIN'S EQUATIONS

Einstein's equations obtained from the variation of the Hamiltonian constraint are a set of ordinary differential equations (ODE's) which may be easily integrated numerically for any convenient choice of variables. Care must be taken to achieve high accuracy since error magnifying subtractions are required in the MSS potential term. Furthermore, the correct solution will not be found unless the Hamiltonian constraint is preserved (which is never guaranteed numerically) [31]. We shall assume that a sufficiently accurate numerical evolution will reveal the true features of Mixmaster dynamics. The β_\pm -plane projection of a generic trajectory, in the direction toward the singularity ($\Omega \rightarrow -\infty$), illustrates the most prominent aspects of Mixmaster dynamics [11, 12, 13]. Epochs of a typical trajectory are shown in Fig. 1.

- (1) Near the maximum of expansion, the MSS potential cannot be neglected anywhere on the trajectory. This leads to bounces off the MSS potential without straight line motion (Kasner) between them.
- (2) As the trajectory scale increases, it consists of a sequence of straight line segments (epochs) joined by bounces off the potential walls.
- (3) The trajectory moves outward along a corner of the potential with segments of successive epochs becoming more and more perpendicular to the potential wall.
- (4) After the trajectory direction crosses the perpendicular so that it no longer moves outward, it travels to a new corner.
- (5) The trajectory is sensitive to the angle between the trajectory and the wall perpendicular. This implies SIC which is one definition of chaos.

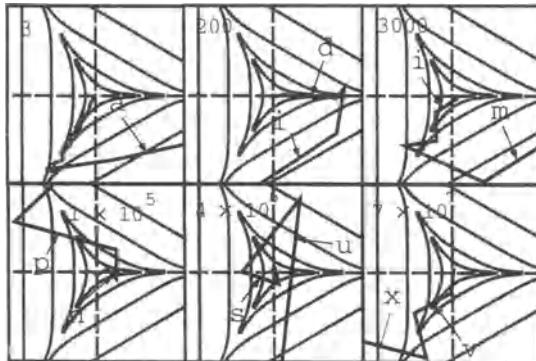


Figure 1. A single Mixmaster trajectory (from [12]) is shown superposed on contours of the MSS potential in the β_{\pm} -plane. The number in the upper left of each frame denotes the scale of the axes. The trajectory (in the approach to the singularity) proceeds from left to right and top to bottom. The letters label the epochs. The same potential contours are shown in each frame. If the contours were scaled to the β_{\pm} axes, only contours which are close to equilateral triangles would appear far from the origin. It is to be noted that the contour $U = 0$ has a fixed location in the β_{\pm} -plane.

Although the trajectory has been displayed in MSS, the same dynamics can be discussed in terms of metric oscillations: As the singularity is approached, there is one expanding direction (say a) and 2 collapsing ones (b, c) at any instant. Two oscillate ($a \leftrightarrow b$) while c decreases monotonically. With each oscillation, the slope of c decreases until it becomes the expanding direction (say $c \leftrightarrow a$) while a now decreases monotonically. A portion of the trajectory in Fig. 1 is shown in terms of the metric scale factors in Fig. 2.

Note that the Mixmaster universe is not physically significant as a cosmological model [2] although its dynamics may characterize a generic inhomogeneous singularity [35]. In terms of comoving proper time, $t_{Hubble}/t_{Planck} \approx 10^{61}$ which (no matter how the correspondence is made) gives $|\Delta\Omega| < 10^3$. But a typical numerical evolution consists

of approximately 30 epochs corresponding to $|\Delta\Omega| \approx 10^8$. This means that independent of details of scaling most bounces occur before the Planck time where classical General Relativity is invalid.

3. THE BKL APPROXIMATE DYNAMICS

Although Einstein's equations for the Mixmaster universe cannot be solved exactly, the dominant features of the dynamics can be studied through the approximate solution first given by BKL (see for example [9]). The assumptions made in this approximation are [34]:

- (1) The MSS potential is replaced by

$$\mathcal{U} = e^{4\alpha} + e^{4\zeta} + e^{4\gamma}. \quad (9)$$

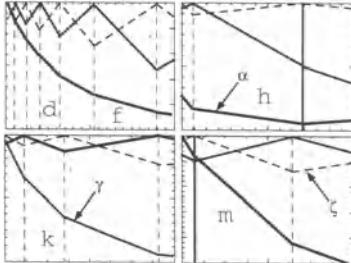


Figure 2. Behavior of the logarithms of metric components α , β , and γ for part of the trajectory of Fig. 1 (adapted from [13]). Again the evolution toward the singularity proceeds from left to right and top to bottom. The scales are different on each frame in correspondence to the increase in scale in Fig. 1. The symbol used for each metric component is consistent from frame to frame and indicated where convenient. The roman letters are again epoch labels and correspond to those in Fig. 1.

Thus the true potential is replaced by one in which the equipotentials are equilateral triangles.

(2) Bounces occur when α , ζ , or $\gamma = 0$ at a fixed Ω (i.e. the bounces are instantaneous in Ω). This means that the triangular equipotentials on which the bounces occur can be labeled by Ω and the geometry of the triangles determined. This is shown in Fig. 3.

(3) Between bounces, the solution is a Kasner universe.

We shall see that the BKL approximate dynamics is then an infinite sequence of Kasner epochs and can thus be expressed as a map for the discrete sequence of parameters u_n describing the successive epochs. This is the era-epoch or u -map [41].

A single bounce off the MSS potential wall is called an epoch change and is equivalent to a Bianchi Type II cosmology. If the scale factor a is expanding in the approach to the singularity, the $\alpha = 0$ wall will dominate. Calculation of the scattering off the wall relates the two asymptotic Kasner epochs. Given $\alpha = \Omega - 2\beta_+$, we define $\omega = 2\Omega - \beta_+$ with conjugate momenta $p_\Omega = p_\alpha + 2p_\omega$, $p_+ = -2p_\alpha - p_\omega$. Then the vanishing of the Hamiltonian constraint means that

$$p_\omega^2 = p_\alpha^2 + \frac{1}{3}p_-^2 + \frac{\xi}{3}e^{4\alpha}. \quad (10)$$

Since the Kasner indices p_i are proportional to $\dot{\alpha}$, $\dot{\zeta}$, and $\dot{\gamma}$ (where $\cdot = d/d\tau$), $p_\alpha \rightarrow -p_\alpha$ with p_ω and p_- constant are sufficient to yield the transformation laws for the indices [9]:

$$p_{1'} = \frac{-p_1}{2p_1 + 1}; \quad p_{2'} = \frac{2p_1 + p_2}{2p_1 + 1}; \quad p_{3'} = \frac{2p_1 + p_3}{2p_1 + 1}. \quad (11)$$

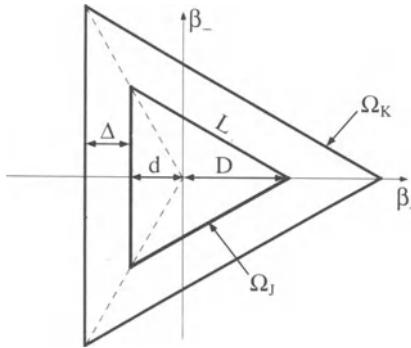


Figure 3. Self-similar triangles in the anisotropy plane. Triangles are labeled by Ω . The side length is $L = \sqrt{3}\Omega$. The distance from the center to the apex is $D = \Omega$ and that to the wall is $d = \Omega/2$. The distance between the walls of triangles labeled by Ω_K and Ω_J is $\Delta = (\Omega_K - \Omega_J)/2$. (For convenience, we treat Ω as > 0 .)

Note that if $p_1 < 0$ and $p_2 < p_3$, then $p_{2'} < 0$ with $p_{1'} < p_{3'}$. Subsequent epochs thus correspond to oscillations between an expanding metric scale factor a and a collapsing one b and vice versa. Reparametrizing the indices in terms of u' gives $u' = u - 1$. In MSS, the trajectory orientation becomes closer to the wall perpendicular as it moves out the corner.

This process cannot continue indefinitely. Eventually, the system undergoes an era change. The Kasner parametrization requires $u \geq 1$. If $1 \leq u \leq 2$, define $u' = (u - 1)^{-1}$ to restore original range. This leads to a change in ordering of the Kasner indices such that if initially $p_1 < p_2 < p_3$ then

$$p_1 < p_2 < p_3 \rightarrow p_2 < p_1 < p_3 \rightarrow p_2 < p_3 < p_1. \quad (12)$$

The expanding a at the end of this era becomes the monotonically decreasing scale factor in the next era, etc. In MSS, the trajectory is on the “wrong” side of the wall perpendicular and moves to a new corner. Clearly, whether or not the MSS trajectory is on the wrong side of the wall perpendicular is an example of SIC. This appears in the map for u through the subtraction in the denominator when $1 \leq u \leq 2$. Such a subtraction is, of course, sensitive to the details of the fractional part of u . The u -map (or epoch-era map) can be summarized as

$$u_{n+1} = \begin{cases} u_n - 1 & , \quad u_n > 2 \\ \frac{1}{u_n - 1} & , \quad 1 \leq u_n \leq 2 \end{cases} \quad (13)$$

The invariant probability that the n th iteration of the map will be between u and $u + du$ is easily computed to be [11]

$$\mu(u)du = \frac{1}{u}du \quad (14)$$

for $1 \leq u < \infty$. But $\mu(u)$ is not normalizable. To regain the concept of invariant probability, a cutoff such as u_{max} , the maximum value of u on a given trajectory, must be imposed.

The Liapunov exponent (LE) measures the divergence of trajectories from nearby initial data [37]. A system described by M first order equations will have M LE’s. (If these equations are derivable from a conserved Hamiltonian, the LE’s will sum to zero.) At least one positive LE is said to characterize a system with chaos. For a discrete realization of a map, the LE σ is given by

$$\sigma(x_0) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \left| \frac{df_N}{dx_0} \right| \quad (15)$$

for two orbits of a map $f(x)$ with initial conditions x_0 and $x'_0 = x_0 + dx_0$. Alternatively, $\sigma = \int_0^1 \ln \left| \frac{df}{dx} \right| \mu(x) dx$ (for $0 \leq x \leq 1$). For the u -map,

$$\sigma = \frac{\pi^2}{6(\ln 2)(\ln u_{max})}. \quad (16)$$

This is a cutoff dependent positive value. For any given trajectory characterized by u_{max} , there will be a positive LE. Yet, the u -map will eventually reach all values of u requiring infinite cutoff. In this case, we find an LE of zero. It is not clear therefore whether the u -map should be regarded to be chaotic although it displays SIC. (For want of a better term, let us call it “marginally chaotic.”) SIC occurs only during the era change. In MSS, the trajectory angle near the wall perpendicular displays sensitivity to continuing out the original corner or moving to a new corner. There is no SIC in epoch changes.

The behavior of the era changes may be summarized by the Gauss map [1]

$$u_{N+1} = \frac{1}{u_N - [u_N]}. \quad (17)$$

with invariant probability

$$\mu(u) = \frac{1}{u(1+u)\ln 2} \quad (18)$$

and LE

$$\sigma = \frac{\pi^2}{6(\ln 2)^2}. \quad (19)$$

This map is definitely chaotic [1].

To relate the BKL u -map to the numerically evolved trajectories we introduce MSS proper time which mimics the discreteness of the BKL approximation [11]. Define this new time variable λ by

$$d\lambda = |p_0^2 - p_+^2 - p_-^2|^{1/2} d\tau. \quad (20)$$

Note that the change in λ , $\Delta\lambda$, equals zero during a Kasner epoch since the quantity inside the square root is proportional to the Hamiltonian constraint for Bianchi Type I. During any bounce, it is easy to show [11] (using the same scattering calculation that leads to the u -map) that

$$\Delta\lambda = \frac{\pi}{4\sqrt{3}}. \quad (21)$$

This time variable allows us to draw an analogy: The sequence $\{n, u_n\}$ is equivalent to the sequence $\{\lambda_n, \theta_n\}$ where θ_n is the MSS trajectory angle during the n th epoch. (The relationship between θ and u and the map for θ are well-known [53, 15, 12].)

For a system of ODE's, the LE's are (for $J = 1, \dots, M$ for $M/2$ degrees of freedom) [37]

$$\sigma_J(\vec{x}_0, \Delta\vec{x}) = \lim_{\substack{t \rightarrow \infty \\ d(0) \rightarrow 0}} \frac{1}{t} \ln \frac{d(\vec{x}_0, t)}{d(\vec{x}_0, 0)} \quad (22)$$

where $d(\vec{x}, t) = \|\Delta\vec{x}(\vec{x}_0, t)\|$ is the distance at time t between two trajectories with initial conditions \vec{x}_0 and $\vec{x}_0 + \Delta\vec{x}$. The LE's computed from Einstein's equations for Ω [16] or τ -time [31] are zero. However, a $\text{LE} > 0$ is found for time variables $T \approx \ln \Omega$ [49] and λ [11]. The difference is understood to be a consequence of exponentially increasing epoch lengths in Ω or τ -time (compared to T or λ -time) [11, 25, 34].

4. THE BOUNCE MAP

The u -map is not invertible since $[u_N]$ is lost and thus cannot describe the time reversibility of Einstein's equations [17]. To regain the time reversibility, we consider a higher dimensional map (an analog of the Poincaré map [37]) known to BKL in the Chernoff-Barrow [19] parametrization. They use 4 variables, but one is just Λ (relating Ω and τ). The remaining map parameters are u , v , and Ω . We shall call the epoch to epoch transformation of $\{u, v, \Omega, \Lambda\}$ the bounce map since it can be shown to be equivalent to sections of the full dynamics at the bounce points [9, 19].

The u, v -map which is the u, v part of the bounce map (in the collapse direction) is given by [13]

$$(u_{n+1}, v_{n+1}) = \begin{cases} (u_n - 1, v_n + 1) & , \quad 2 \leq u_n \leq \infty \\ \left(\frac{1}{u_n - 1}, \frac{1}{v_n} + 1\right) & , \quad 1 \leq u_n \leq 2 \end{cases} \quad (23)$$

while the extended Gauss version relating the eras is [17] (for $1 \leq u_N, v_N < \infty$)

$$(u_{N+1}, v_{N+1}) = \left(\frac{1}{u_N - [u_N]}, \frac{1}{v_N} + [u_N] \right). \quad (24)$$

The inverse map is obtained by the interchange of u and v since $[v_{N+1}] = [u_N]$. It is easy to see that the maps for u and v are the inverses of each other. That is, if $\{u_{n+1}, v_{n+1}\} = \{f(u_n), g(v_n)\}$ then $\{u_n, v_n\} = \{g(u_{n+1}), f(v_{n+1})\}$. There is some arbitrariness in the association of u and v with particular Kasner segments. Consistency of assignments can be checked with numerical experiments [11, 13, 14]. The u , v -map has two LE's—one is the same as that for the u -map while the other vanishes. This means u (v) is marginally chaotic (regular) in the collapse direction and regular (marginally chaotic) in the expansion direction [13].

At any bounce, there is a fixed value of Ω and one among α , ζ , and γ vanishes. Consider an era evolving in the 120° corner. The non-vanishing α or ζ measures the distance to the corner while γ measures the distance to the wall opposite the corner. This is illustrated in Fig. 4. Say $\zeta_0 = 0$ at Ω_0 . Since $\alpha + \zeta + \gamma = 3\Omega$, define v by the parametrization [19, 13]

$$\alpha_0 = \frac{3\Omega_0}{1 + u(v-1)}; \quad \gamma_0 = \frac{3\Omega_0 u(v-1)}{1 + u(v-1)}. \quad (25)$$

The Kasner indices in terms of u are Ω derivatives of α , ζ , and γ :

$$\frac{d\alpha}{d\Omega} = \frac{3(1-u)}{u^2 - u + 1}, \quad \frac{d\zeta}{d\Omega} = \frac{3u}{u^2 - u + 1}, \quad \frac{d\gamma}{d\Omega} = \frac{3(u^2 - u)}{u^2 - u + 1}. \quad (26)$$

Note that u is offset by 1 epoch from the usual definitions [9].

Evolve to the next bounce at Ω_1 where $\alpha_1 = 0$. We find [9, 19, 13, 14]

$$\zeta_1 = \frac{3\Omega_0 u}{(u-1)\{1+u(v-1)\}}; \quad \gamma_1 = \frac{3\Omega_0 uv}{1+u(v-1)} \quad (27)$$

with

$$\Delta\Omega = \frac{3\Omega_0(u^2 - u + 1)}{(u-1)\{1+u(v-1)\}} \quad (28)$$

where $\Delta\Omega = \Omega_1 - \Omega_0$. Solve to obtain u and v as

$$u = \frac{\zeta_1}{\zeta_1 - \alpha_0}; \quad v = \frac{\gamma_1}{\gamma_1 - \gamma_0}. \quad (29)$$

To see that u and v should satisfy the same maps, we compute the ratio of γ to α or ζ for a single wall. For the bounces at Ω_0 and Ω_1 we find [13]

$$\frac{\gamma_0}{\alpha_0} = (v-1)u; \quad \frac{\gamma_1}{\zeta_1} = v(u-1). \quad (30)$$

which according to the u,v -map is the appropriate update. Note the essentially symmetric appearance of u and v in this ratio. At the end of an era, (say at Ω_1) the roles of ζ and γ will interchange so that ζ now measures the distance to the wall opposite the new era's corner. If u' , v' describe the first epoch of the new era, the u,v -map correctly gives

$$\frac{\zeta_1}{\gamma_1} = (v'-1)u' = \frac{1}{v(u-1)}. \quad (31)$$

To further elucidate the nature of the era change in both the forward and reverse time directions, we introduce the era scale parameter [13]. The parameter κ defined by

$$\kappa = \frac{3\Omega_0 u}{1 + u(v - 1)} \quad (32)$$

is invariant from epoch to epoch of a given era. Substitution for u, v gives

$$\kappa = \gamma_1 - \gamma_0 = \frac{\alpha_0 \zeta_1}{\zeta_1 - \alpha_0}. \quad (33)$$

(The two expressions for κ are identically equal by virtue of the equations governing the Kasner indices.) We see that the value of κ associated with a given era defines a

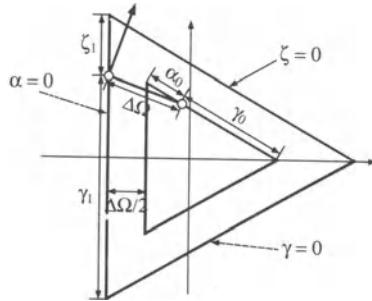


Figure 4. Bounce parameters. For the triangles shown in Fig. 3 (now labeled by Ω_0 and Ω_1), the dark solid line is a Kasner segment. The bounce points divide the $\zeta_0 = 0$ and $\alpha_1 = 0$ walls into α_0 , γ_0 and ζ_1 , γ_1 respectively (where for clarity the factor of $\sqrt{3}$ that divides all the bounce parameters has been suppressed). The segment length $\Delta\Omega$ and wall separation $\Delta\Omega/2$ are consequences of the fact that the Kasner solution requires the system point to have unit velocity while the wall expands at an equivalent velocity of $\frac{1}{2}$ [54]. The walls are labeled by the scale factor logarithm that vanishes there.

fixed scale for that era. In terms of κ ,

$$u = \frac{\kappa}{\alpha_0}, \quad v = \frac{\gamma_1}{\kappa}. \quad (34)$$

In the collapse direction, the equipotential triangles increase in size. Since κ is fixed, u decreases (while v increases). The trajectory orientation moves closer to the wall perpendicular. When u falls below unity, the era must change. In the expansion direction, the triangles decrease in size so that v decreases (while u increases). When v falls below unity, κ cannot fit in the equipotential triangle so the era must change. Thus u (v) leads to the SIC associated with the era change in the collapse (expansion) direction. For the next era,

$$\kappa_{N+1} = u_{N+1} v_N \kappa_N \quad (35)$$

where u_{N+1} is the starting value of u in the $N + 1$ st era and v_N is the final value of v in the N th era (in the collapse direction). Thus we see that the BKL approximate dynamics can describe and explain the observed numerical evolution.

5. DISCUSSION

What could go wrong with the BKL description of Mixmaster dynamics? Are significant features of Mixmaster dynamics absent from the BKL approximation? First, could numerical errors have produced a wrong description of Mixmaster dynamics? This seems unlikely since (1) the numerical trajectories can be run backwards [13] and (2) excellent agreement for u -map and bounce-map parameters from numerical trajectories has been found [51, 11, 12, 13, 14]. (See Table 2 [14].)

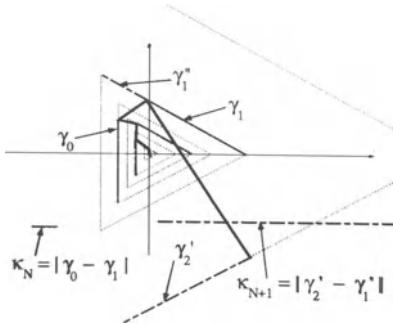


Figure 5. The change of an era in terms of its parameter κ [13]. Self-similar triangles are constructed from the bounce points of an actual trajectory [11] (solid dark line). γ measures the distance from the bounce point to the wall not involved in that era's corner. κ is constant during an era. Toward the singularity, the increasing scale causes the next bounce point to be pulled to the other side of the wall perpendicular. During expansion, the decreasing scale leaves no room for the next γ . We have treated all γ 's as positive and dropped factors of $\sqrt{3}$.

Second, it is well-known that the BKL approximation is invalid if there is no Kasner era. This is true only near the maximum of expansion or when the trajectory is close to a corner ray of the MSS potential where the difference between the true and approximate potential cannot be neglected. Near the maximum of expansion, the evolution is qualitatively like the BKL approximation. Trajectories which pass through the maximum of expansion are very sensitive to initial conditions [10]. Close to the corner ray, we have the so-called anomalous behavior [9, 48]. When the MSS trajectory angle is within

$$\theta = \frac{\ln 2}{4\sqrt{3} |\beta_+|} \quad (36)$$

Table 2. Evaluated and predicted BKL parameters: u , v , and κ are evaluated directly from the numerical trajectories. The predictions are based on the u,v -map and the era to era evolution rule for κ . Values of u and v indicated by the same symbol should be multiplied together to evolve κ . (Predicted values for adding and subtracting 1 are not shown.) Better agreement for u compared to v or κ is due to the fact that the u -map requires only the BKL assumption that the interbounce dynamics be Kasner while the rest of the bounce map requires all the BKL assumptions.

epoch	u	u_{pred}	v	v_{pred}	κ	κ_{pred}
a	7.229326		2.03968		-68.250	
b	6.231434		3.11097		-69.036	
c	5.231434		4.14907		-69.680	
d	4.231435		5.16301		-70.150	
e	3.231436		6.16326		-70.407	
f	2.231437		7.16359		-70.395	
g	1.231441		8.18506*		-70.012	
h	4.320915*	4.3208	1.12093	1.1222	-2466.6	- 2476.0
i	3.320981		2.11641		-2464.2	
j	2.320982		3.11380		-2461.2	
k	1.320984		4.11352†		-2457.9	
l	3.115458†	3.1153	1.24294	1.2431	-31472.	- 31499.
m	2.115468		2.24244		-31468.	
n	1.115469		3.24229‡		-31462.	
o	8.660378‡	8.6603	1.30840	1.3084	-883330.	- 883610.

of the (for convenience) 0° corner ray,

$$\mathcal{U} \propto e^{4\Omega+4\beta_+} \beta_-^2 \quad (37)$$

so the behavior is qualitatively different from that assumed in the BKL approximation. The requirement for anomalous behavior is equivalent to $u \geq \sqrt{3}/\theta$ [11]. However, in a sequence of eras $\{u_N\}$ the probability of $u > u_0 \approx 1/u_0^2$ [1]. Since $|\beta_+| \approx |\Omega|$ and $|\Omega| \rightarrow \infty$ as the singularity is approached, the anomalous behavior becomes less and less likely. Even if anomalous behavior should occur, eventually, the approximate dynamics will be regained [9]. (The solution for anomalous behavior along the positive β_+ axis can be expressed in terms of Bessel functions of zero order

$$Z_0 \left[\frac{2\sqrt{3}}{p_\Omega - p_+} \exp(2\Omega + 2\beta_+) \right] \quad (38)$$

which becomes the Kasner solution in the small argument limit. This occurs because $p_- \neq 0$ although it can be initially very small. As p_- and β_- grow in magnitude, the β_- degree of freedom shares more of the “energy” represented through the Hamiltonian constraint by p_Ω^2 . Eventually, the Ω -dependent terms dominate the argument of the Bessel function driving it down since $\Omega \rightarrow -\infty$ in the approach to the singularity.)

It should still be shown rigorously that all the BKL assumptions improve as the singularity is approached. Heuristically, the improvement occurs because (1) \mathcal{U} away from the bounce evolves to become an ever smaller fraction of \mathcal{U} at the bounce so that the Kasner interbounce behavior becomes a better approximation; (2) the bounce duration in Ω -time becomes an ever smaller fraction of the epoch duration; and (3) the actual values of α , ζ , and γ at the bounce are ever smaller fractions of typical interbounce values. The validity of the BKL approximation must be considered in combination with a study of how the accumulation of errors in any numerical evolution of the Mixmaster Einstein equations might eventually cause noticeable deviation from the true solution. A possible method of attack for this combined problem is to compute all the BKL parameters from the numerical solution as Einstein's equations are integrated. An algorithm exists to do this [14].

There are also features of Mixmaster dynamics associated with special values of u that should be explored [9, 43]: (1) If u is rational, the map ends at $u = \infty$ corresponding to the $(1, 0, 0)$ Kasner which is a trajectory that heads straight out the corner (no more bounces). (2) The Gauss map has fixed points $u_{N+k} = u_N$. For example, $u = (1 + \sqrt{5})/2$ is a fixed point for $k = 1$.

Finally, we come to the question of whether or not the Mixmaster dynamics is chaotic. Numerical studies suggest that the BKL approximate solution accurately describes Mixmaster dynamics as $\Omega \rightarrow -\infty$ so that the features of the dynamics are well-understood [11, 12, 13]. Questions about the dynamics can then be answered in terms of properties of the bounce map. The substantive statement that can be made is that there is no reason to believe that the Mixmaster oscillations will either end or change their character as $\Omega \rightarrow -\infty$. This issue could be addressed by a rigorous proof of the validity of the BKL approximation. Of course, it is also in principle possible to approach the question from the opposite viewpoint to decide whether or not Mixmaster dynamics is integrable as has been suggested from evidence relating to the Painlevé criterion [20, 21]. Equivalently, one may search for non-trivial constants of the motion [27]. So far, none have been found. It has been shown that there are no constants of the motion up to quadratic order in the canonical momenta [55]. The existence of one such constant may be plausible since (in the approach to the singularity) two of the three BKL parameters u , v , and Ω are not sensitive to initial conditions.

Ultimately, the question of whether or not Mixmaster dynamics is chaotic may be merely semantic. For example, the LE for the u -map is a trajectory dependent positive number with limiting value zero. The standard definitions of chaos do not apply to Mixmaster dynamics without revision. Yet any revised criterion might yield a result which depends on how heavily it weights those aspects of the dynamics which are sensitive to initial conditions compared to those which are not. If the Mixmaster equations cannot be shown to be integrable, we may have to be satisfied with the term "marginal chaos" to describe the dynamics.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grants PHY-8904847, PHY-9107162, and PHY-930559 to Oakland University. The author is grateful for hospitality provided by the Institute for Geophysics and Planetary Physics at Lawrence Livermore National Laboratory. Computations were performed with the facilities of the National Center for Supercomputing Applications at the University of Illinois.

REFERENCES

- [1] Barrow, J.D., 1982, *Phys. Rep.*, **85**, 1.
- [2] Barrow, J.D., 1984, *Classical General Relativity*, (Cambridge: Cambridge University), p 25.
- [3] Barrow, J.D. and Tipler, F., 1979, *Phys. Rep.*, **56**, 372.
- [4] Belinskii, V.A. and Khalatnikov, I.M., 1969a, *Sov. Phys. JETP*, **29**, 911.
- [5] Belinskii, V.A. and Khalatnikov, I.M., 1969b, *Sov. Phys. JETP*, **30**, 1174.
- [6] Belinskii, V.A. and Khalatnikov, I.M., 1971, *Sov. Phys. JETP*, **32**, 169.
- [7] Belinskii, V.A., Khalatnikov, I.M. and Lifshitz, E.M., 1970, *Adv. Phys.*, **19**, 525.
- [8] Belinskii, V.A., Khalatnikov, I.M. and Lifshitz, E.M., 1982, *Adv. Phys.*, **31**, 639.
- [9] Belinskii, V.A., Lifshitz, E. M., and Khalatnikov, E. M. , 1971, *Sov. Phys. Usp.*, **13**, 745.
- [10] Berger, B.K., 1990, *Class. Quantum Grav.*, **7**, 203.
- [11] Berger, B.K., 1991, *Gen. Rel. Grav.*, **23**, 1385.
- [12] Berger, B.K., 1992, *Proceedings of the 4th Canadian Conference on General Relativity and Relativistic Astrophysics*, (Singapore: World Scientific).
- [13] Berger, B.K., 1993a, *Phys. Rev. D*, **47**, 3222.
- [14] Berger, B.K., 1993b, "How to determine approximate Mixmaster parameters from numerical evolution of Einstein's equations," preprint.
- [15] Bugalho, M.H., Rica da Silva, A. and Sousa Ramos, J., 1986, *Gen. Rel. Grav.*, **18**, 1263.
- [16] Burd, A.B., Buric, N., and Ellis, G.F.R. , 1990, *Gen. Rel. Grav.*, **22**, 349.
- [17] Burd, A.B., Buric, N., and Tavakol, R.K., 1991, *Class. Quantum Grav.*, **8**, 123.
- [18] Burd, A. and Tavakol, R., 1993, *Phys. Rev. D*, **47**, 5336.
- [19] Chernoff, D.F., Barrow, J.D., 1983, *Phys. Rev. Lett.*, **50**, 134.
- [20] Contopoulos, G., Grammaticos, B. and Ramani, A., 1993, "Painlevé analysis for the Mixmaster Universe model," preprint.
- [21] Cosakis, S. and Leach, P.G.L., 1993, "Painlevé analysis of the Mixmaster universe," preprint.
- [22] Demaret, J. and De Rop, Y., 1993, *Phys. Lett.*, **B299**, 223.
- [23] Elskens, Y., 1983, *Phys. Rev. D*, **28**, 1033.
- [24] Ferraz, K. and Francisco, G., 1992, *Phys. Rev. D*, **45**, 1158.
- [25] Ferraz, K., Francisco, G. and Matsas, G.E.A., 1991, *Phys. Lett.*, **156A**, 407.
- [26] Francisco, G. and Matsas, G.E.A., 1988, *Gen. Rel. Grav.*, **20**, 1047.
- [27] Grubisic, B. and Moncrief, V., 1993, "The Mixmaster spacetime, Geroch's transformation and constants of motion," preprint.
- [28] Gutzwiller, M.C., 1990, *Chaos in Classical and Quantum Mechanics*, (New York: Springer-Verlag).
- [29] Halpern, P., 1987, *Gen. Rel. Grav.*, **19**, 73.
- [30] Higuchi, A. and Wald, R., 1993, private communication.
- [31] Hobill, D., Bernstein, D., Welge, M. and Simkins, D., 1991, *Class. Quantum Grav.*, **8**, 1155.

- [32] Ishihara, H., 1985, *Prog. Theor. Phys.*, **74**, 490.
- [33] Jantzen, R.T., 1984, *Cosmology of the Early Universe*, (Singapore: World Scientific), p 233.
- [34] Khalatnikov, I.M., Lifshitz, E.M., Khanin, K.M., Shchur, L.N., and Sinai, Ya. G., 1985, *J. Stat. Phys.*, **38**, 97.
- [35] Khalatnikov, I.M. and Pokrovski, V.L., 1972, *Magic without Magic*, (San Francisco: Freeman), p 441.
- [36] Kodama, H., 1988, *Prog. Theor. Phys.*, **80**, 1024.
- [37] Lichtenberg, A.J. and Lieberman, M.A., 1983, *Regular and Stochastic Motion*, (New York: Springer-Verlag).
- [38] Lifshitz, E.M. and Khalatnikov, I.M., 1964, *Sov. Phys. Usp.*, **6**, 495.
- [39] Lin, X. and Wald, R., 1989, *Phys. Rev. D*, **40**, 3280.
- [40] Lin, X. and Wald, R.M., 1990, *Phys. Rev. D*, **41**, 2444.
- [41] Ma, P.K.-H. and Wainwright, J., 1991, *Proceedings of the Third Hungarian Relativity Workshop*.
- [42] MacCallum, M., 1979, *General Relativity, an Einstein Centenary Survey*, (Cambridge: Cambridge University).
- [43] Misner, C.W., 1969a, *Phys. Rev. Lett.*, **22**, 1071.
- [44] Misner, C.W., 1969b, *Phys. Rev.*, **186**, 1328.
- [45] Misner, C.W., 1970, *Relativity*, (New York: Plenum), p 55.
- [46] Misner, C.W., 1972, *Magic without Magic*, (San Francisco: Freeman), p 441.
- [47] Misner, C.W., Thorne, K.S. and Wheeler, J.A., 1973, *Gravitation*, (San Francisco: Freeman), 810–813.
- [48] Moser, A.R., Matzner, R.A., and Ryan, M.P. Jr., 1973, *Ann. Phys., NY*, **79**, 558.
- [49] Pullin, J., 1991, *SILARG VII Relativity and Gravitation: Classical and Quantum*, (Singapore: World Scientific).
- [50] Rugh, S.E., 1990, Cand. Scient. Thesis, Niels Bohr Inst.
- [51] Rugh, S.E. and Jones, B.J.T., 1990, *Phys. Lett.*, **A147**, 353.
- [52] Ryan, M.P. Jr. and Shepley, L.C., 1975, *Homogeneous Relativistic Cosmologies*, (Princeton: Princeton University).
- [53] Ryan, M.P. Jr., 1971, *Ann. Phys., NY*, **65**, 506.
- [54] Ryan, M.P. Jr., 1972, *Hamiltonian Cosmology*, (New York: Springer- Verlag).
- [55] Schleich, K., 1993, private communication.
- [56] Taub, A., 1951, *Ann. Math.*, **53**, 472.
- [57] Wainwright, J. and Hsu, L., 1989, *Class. Quantum Grav.*, **6**, 1409.

HOW CAN YOU TELL IF THE BIANCHI IX MODELS ARE CHAOTIC?

Adrian Burd ¹

Dept. of Oceanography, Dalhousie University
Halifax, Nova Scotia, Canada B3H 4J1

Abstract. In this article the various techniques that have been used to show that the Bianchi IX models are chaotic are reviewed. It is shown that all the methods used so far contain certain pitfalls and none provide a conclusive, rigorous demonstration that these models exhibit chaotic behaviour; this is due in part to the nature of the Bianchi IX system. Some suggestions for possible directions for future research are given.

1. INTRODUCTION

Chaos theory, or non-linear dynamical systems theory to give it its less glamorous title, has been around for many years now. Some of what passes under this subject heading would not be unfamiliar to the likes of Poincaré and Hadamard. In fact much of the terminology would be familiar; Poincaré (1899) for example coined the term homoclinic point. Whilst mathematicians have concerned themselves with these things since the late nineteenth century, physicists, engineers, atmospheric scientists and even oceanographers have only recently become enamoured with the subject.

In spite of the high degree of non-linearity exhibited by the Einstein equations, relativists have generally not become involved in the theory of chaos and its possible application in their own field. Until recently there was one exception and that was the study of the evolution of the Bianchi IX model (see e.g. Belinski *et al* 1970). The behaviour of this model was shown to be stochastic by Belinski *et al* some time ago and this was confirmed by Barrow (1982) who applied dynamical systems theory to the system. Since then, a controversy has grown up concerning whether the Bianchi IX

¹e-mail: aburd@open.dal.ca

model is, or is not, chaotic (see e.g. Francisco and Matsas 1988). This controversy has been fuelled to a certain extent by the use of standard tools from dynamical systems theory in situations where they are not really applicable.

Here I give a review of the various methods that have been used to demonstrate the chaotic behaviour of the Bianchi IX model and highlight their pitfalls and suggest directions for further research.

2. LYAPUNOV EXPONENTS

Barrow (1982) was the first to explicitly show, using dynamical systems theory, that the Bianchi IX model was chaotic (see also Khalatnikov *et al* 1985). This was done by using the Gauss map derived from the flow and showing that this map had a non zero Kolmogorov entropy. This in turn implies that the system has at least one positive Lyapunov exponent (LE). This result was later supported by Zardecki (1983) who performed a series of numerical experiments on the full flow and showed that the principal LE was positive. The Lyapunov exponent is a measure of the mean exponential rate of divergence of neighbouring trajectories and a positive value indicates that the dynamical system is chaotic.

A few years later Francisco and Matsas (1988) performed similar numerical experiments but found surprisingly that as the singularity was approached, the principal LE tended asymptotically to zero implying that the system is not chaotic. This was not only in contradiction to the previous numerical result but also with the analytic result obtained by Barrow. This was explained by Francisco and Matsas as being due to the time variable that they used in their numerical simulations and was confirmed later by others (Burd *et al* (1990), Hobill *et al* (1991), Rugh and Jones (1990), Berger (1990, 1991)).

Let us return to the beginning of this story and look at it in a bit more detail. The Bianchi IX metric can be written in the form

$$ds^2 = -dt^2 + \eta_{ik} \omega^i \omega^k \quad (1)$$

where $\eta_{ik} = \text{diag}(a^2(t), b^2(t), c^2(t))$. The vacuum field equations can then be written in the form

$$\begin{aligned} (\dot{abc}) &= -\frac{1}{2abc} [a^4 - (b^2 - c^2)^2] \\ (\dot{bca}) &= -\frac{1}{2abc} [b^4 - (c^2 - a^2)^2] \\ (\dot{cba}) &= -\frac{1}{2abc} [c^4 - (a^2 - b^2)^2] \\ 0 &= \frac{\ddot{a}}{a} + \frac{\ddot{b}}{b} + \frac{\ddot{c}}{c} \end{aligned}$$

(for more details see the article by Hobill in this volume). In the case when the right hand sides are zero, these equations can be solved giving the Kasner cosmological model. Belinski *et al* were able to show that the evolution of the Bianchi IX model towards the singularity was well approximated by an infinite sequence of such Kasner models (Belinskii *et al* 1970). The Kasner solution exhibits a power law behaviour with $a(t) = t^{p_1}$, $b(t) = t^{p_2}$ and $c(t) = t^{p_3}$ with the constraints that $\sum p_i = \sum p_i^2 = 1$; the solution is

thus a one-parameter solution and the Kasner indices, p_i , can be expressed in terms of the Kasner parameter u via

$$p_1 = \frac{-u}{1+u+u^2}, \quad p_2 = \frac{1+u}{1+u+u^2}, \quad p_3 = \frac{u(1+u)}{1+u+u^2}. \quad (2)$$

The Gauss map is a discrete mapping describing the evolution of the Bianchi IX model in terms of successive Kasner regimes and is given by (for more details, see the contribution by Berger in this volume)

$$x_{n+1} \equiv u_{n+1} - [u_{n+1}] = x_n^{-1} - [x_n^{-1}] \quad (3)$$

where $[\dots]$ indicates that one should take the integer part of the bracketed quantity.

Barrow has made an extensive study of this map (Barrow 1982, see also Mayer 1987 and Khalatnikov *et al* 1985). In particular, he was able to give an explicit expression for the Kolmogorov entropy, h , associated with the Gauss map

$$h = \frac{\pi^2}{6(\ln 2)^2}. \quad (4)$$

The Kolmogorov entropy is a measure of the information lost as the system evolves and a theorem due to Pesin (1977) relates h with the Lyapunov exponents; for an N -dimensional system

$$h = \sum_{i=1}^N \lambda_i^+ \quad (5)$$

where λ_i^+ are the positive LEs. Thus a non-zero value for h implies that at least one of the LEs must be positive and hence the underlying system will exhibit chaotic behaviour. So, one can claim that at least the discretized approximation to the Bianchi IX cosmology is chaotic.

This raises the following question: since a discretized approximation to the system exhibits chaos, does this imply that the continuous system also exhibits chaotic behaviour? This is not necessarily a trivial question since it is known that the behaviour of discrete dynamical systems can be more complex than that of the corresponding continuous system.

To analyse the full flow it is necessary to look at the dynamical system formed by the ODEs representing the Bianchi IX model. However, apart from a few special cases such as the Taub model, there are no known analytical solutions to these equations so one must resort to the use of numerical simulation. Zardecki (1983) used a standard ODE solver to analyse the full system for a variety of initial conditions and assumptions on the matter stress-energy tensor. He also calculated the principal LE and showed that it was positive for Bianchi IX thus confirming the analytical results. A few years later Francisco and Matsas (1988) performed a set of similar numerical experiments. Their results showed that the principal LE actually tended to zero asymptotically as the singularity was approached. Similar results were also obtained by Rugh and Jones (1990), Burd *et al* (1990), Hobill *et al* (1991) and Berger (1991). This apparent conflict between the results obtained by Barrow and Zardecki and those obtained later have several contributing factors and I shall try to outline them now.

Firstly, there is a numerical problem. This is not, as one might think, a problem caused by inability of the computer itself to calculate and store numbers to infinite precision, but instead results from the numerical methods used; these are explained in

detail in Hobill *et al* (1991). One contributing factor is that in some cases the initial conditions were chosen in such a way as to violate the positive energy conditions (see also Rugh and Jones 1990). As the models evolved towards towards the singularity, the negative energy densities produced by such a choice force the model to bounce rather than collapse into the singularity. These bounces produced a series of chaotic oscillations which led to a spurious value for the lyapunov exponent.

A more subtle problem arises in the choice of numerical algorithm. Since the system we are dealing with is a constrained Hamiltonian system, it is natural to require that any numerical algorithm satisfies the constraint through the numerical evolution of the system. This is not always done. For example, as shown by Hobill at al., certain predictor-corrector methods do not keep the constraint satisfied. This is because, to initiate the predictor-corrector a certain number of steps are taken using a Runge-Kutta algorithm and unfortunately, even though only three steps are taken using this method, errors arise because the convergence of the solution is not checked. The resulting computer simulation then departs from the Bianchi IX solution.

The next contributing factor is the return of an old problem in general relativity, that of gauge invariance. A variety of different coordinates were used in the numerical experiments listed above. For example, the results by Burd *et al* (1990) were derived using the Ellis-MacCallum system (Ellis and MacCallum 1969) (see also Wainwright 1987), Zardecki (1983) used a system of variables favoured by BKL (1970) whereas Berger's results (1991) use the Hamiltonian formulation (for a discussion of the relationship between various of these systems, see Rugh (1989)). The Lyapunov exponents are defined as a time average over the trajectory so that

$$\lambda_i = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{d_{(i)}(x_0, t)}{d_{(i)}(x_0, 0)} \quad (6)$$

where $d_{(i)}$ is a perturbation along the i^{th} principle direction. Oseledec has shown that this definition gives exponents which are independent of the phase space coordinates (Oseledec 1968).

There are two problems with using this definition in general relativity. Firstly, the chaos that occurs in the Bianchi IX model is similar to that found in the classical stadium problems; the chaos is not "produced" uniformly along the trajectory but instead appears in bursts. Whilst here this can be seen intuitively by thinking of the mixmaster model as a ball bouncing inside a potential, chaos can be produced non-uniformly in other cases such as the Rössler system (Burd *et al* 1991) and the Lorenz system (Palmer 1993). Secondly, by using a "bad" choice of time, we can effectively average out all the contributions to the Lyapunov exponent. This is what happens in the above examples.

We can consider this further by recalling that the LEs (λ_i) for a continuous system are related to the corresponding exponents ($\tilde{\lambda}_i$) for the first return map by

$$\lambda_i = \frac{\tilde{\lambda}_i}{\langle \tau \rangle} \quad (7)$$

where $\langle \tau \rangle$ is the average time between two crossings by the flow of the Poincaré section. To see how a change in time variable can affect the LE, consider the following simple non-invertible map

$$X_{n+1} = 2X_n \pmod{1} \quad (8)$$

which possesses nice ergodic properties. The Lyapunov exponent is then

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \ln |X_n| \quad (9)$$

where t is the “time” it takes for n iterations of the map to take place. If we choose the units of time (i.e. the interval between the n and $n + 1$ iterations) to be uniform and equal to unity then $t = n$ and $\lambda = \ln 2$. But, if we choose our clocks such that $t = \exp(n)$ then $\lambda \rightarrow 0$ as $t \rightarrow \infty$.

This effect can also be seen if we calculate the Short Time Approximation (STA) to the LE, also called the local Lyapunov exponent (Burd *et al* 1990, 1991). Instead of seeing smooth convergence to either a non-zero or zero value, the STA show a series of spikes which correspond to the changes in the Kasner evolution in the Mixmaster model (or to the bounces off the potential walls of the universe point in the Hamiltonian picture). The spikes are separated by regions where λ is zero and these increase in length as one approaches the singularity. Thus one can see that the standard definition of the LE gives a value which will average to zero as one allows time to tend to infinity. (As an aside, local Lyapunov exponents are now being used in analysing the predictive power of various atmospheric and climatic computer models, see e.g. Palmer 1993).

Since one can choose a time variable to suit ones needs, is it possible to choose one such that the Lyapunov exponents are positive and give a value similar to that obtained from the analytic results? Berger (1991) has looked into this and shown that a suitable choice would be a form of Minisuperspace proper time. This time variable remains constant except during the bounces when it changes by the same amount each time. In this sense, this allows the evolution of the flow to mimic the evolution obtained from the discrete map. The Lyapunov exponents calculated in these coordinates converge to positive values but it is unclear at the present how these are related to the LE obtained by Barrow.

It must be clear from the above that the Lyapunov exponent is not necessarily a useful tool for studying the chaotic behaviour in general relativity. One can of course choose a suitable gauge and calculate the values of λ in it, but then the result could be purely gauge. Since the LEs for systems studied in dynamical systems are written implicitly in proper time, it can be argued that if one were to make a choice of particular gauge, choosing cosmological proper time would be the most meaningful one. The problem with such a choice is that for evaluating the standard definition of the Lyapunov exponent one requires an effectively infinite time interval. However, it would be nice to have at ones hands a gauge invariant tool for analysing the chaotic behaviour in general relativity. In what follows we shall look at some possibilities.

3. GEODESICS ON A NEGATIVELY CURVED SPACE

One of the first chaotic dynamical systems that was studied described the behaviour of billiards on a surface with negative curvature; this was the work of Jacques Hadamard in 1898. Hadamard couched his results in the language of differential geometry and depended on the curvature of the surface being negative, but not necessarily constant (for a concise discussion of motion on Riemannian spaces see Appendix 1 of Arnold 1989). It can be shown that geodesic motion on a space of negative curvature is ergodic. This result was used by Chitre (1972) to show that under certain approximations the Bianchi IX model was chaotic (see also the contribution by Misner in this volume).

Chitre starts from the Hamiltonian point of view (see the contribution by Berger in this volume) and extends the phase space to include the time coordinate and thus obtaining the so-called super-Hamiltonian

$$H = \alpha \exp(-3\Omega) \left[-p_\Omega^2 + p_+^2 + p_-^2 + e^{4\Omega} V(\beta_+, \beta_-) \right] \quad (10)$$

and where p_Ω , p_+ , p_- are the momenta canonically conjugate to the variables Ω , β_+ and β_- respectively. The potential is related to the curvature of the space-like slices and is given by

$$V(\beta_+, \beta_-) = \exp(-8\beta_+) + 2 \exp(4\beta_+) \left[\cosh(4\sqrt{3}\beta_-) - 1 \right] - 4 \exp(-2\beta_+) \cosh(2\sqrt{3}\beta_-). \quad (11)$$

The potential has a triangular symmetry and as $\Omega \rightarrow \infty$ the walls of the potential move outwards. Chitre considers the asymptotic form of these walls and is able to choose a new set of coordinates such that asymptotically the walls are stationary; the relevant coordinate transformation is

$$\Omega = e^t \cosh \xi \quad (12)$$

$$\beta_+ = e^t \sinh \xi \cos \phi \quad (13)$$

$$\beta_- = e^t \sinh \xi \sin \phi \quad (14)$$

In these coordinates, the potential walls are described by the equations $\tanh \xi = (-1/2) \sec \phi$ with the remaining two walls obtained by making the replacement $\phi \rightarrow \phi \pm 2\pi/3$. If we also assume that asymptotically $V = 0$ inside the potential walls and infinity outside them, then the Hamiltonian can be written as

$$\tilde{H} = \frac{1}{2} \left(-p_t^2 + p_\xi^2 + \frac{p_\phi^2}{\sinh^2 \xi} \right) \quad (15)$$

If one then projects down onto the two-dimensional space (ξ, ϕ) this Hamiltonian represents a geodesic flow on a Riemannian manifold with metric

$$dl^2 = d\xi^2 + \sinh^2 \xi d\phi^2 \quad (16)$$

and the motion being confined within the potential walls. This metric is that of the Lobachevsky plane and the motion consists of arcs of geodesics on this space. Chitre then uses results given by Hopf (1936) and Hedlund (1939) to show that this motion is chaotic.

The approximations that Chitre uses are essentially those of assuming that the potential walls are infinitely hard and straight and that the potential inside the walls is zero. These mimic the approximations used by BKL in their derivation of the BKL map and so it is in many ways not surprising that one should find the motion to be chaotic.

If one could show that the full flow can be represented by a geodesic flow on a negatively curved space then this would be a gauge invariant way of determining the chaotic behaviour of the Bianchi IX models. This approach was advocated by Pullin (1990) and attempted by Szydłowski and his collaborators (see e.g. Szydłowski and Lapeta 1990) This approach, whilst giving a demonstration of local chaos cannot be used globally over the whole space (as pointed out by Misner several years ago).

The essentials of the approach are that one attempts to use the Maupertius Principle to write the Hamiltonian for the Bianchi IX model as that representing geodesic

motion on a Riemannian space. One then examines the curvature of this space to see if it is negative or not. For a free particle it is easy to see that the Hamiltonian can be thought of as that describing geodesic motion on a Riemannian space. When there is a potential energy involved, the motion can be described as a geodesic flow in a manifold which is conformally related to that for the corresponding free motion. For example, for dynamics represented by a Hamiltonian $H = T + V = h = \text{constant}$ where $T = 1/2(ds/dt)^2$ is the kinetic and V is the potential energy, the motion will be that of a geodesic flow in a manifold having $d\tau = ds/(2(h - U))$ (see e.g. Arnold (1989)).

Szydlowski *et al* employ the above procedure to examine the dynamics of the Bianchi IX system. In this case, (10) can be written

$$H = 1/2G^{AB}p_A p_B + \exp(4\Omega)V(\beta_+, \beta_-) \quad (17)$$

where $G^{AB} = \text{diag}(-1, 1, 1)$ and A and B run through the symbols $+$, $-$, Ω . The dynamics can thus be transformed to a geodesic motion on a conformally related space, the metric of which can be written

$$\tilde{G}_{AB} = \pm|V|G_{AB}. \quad (18)$$

The scalar curvature of this new manifold is given by

$$\tilde{R} = -\frac{1}{V} \left\{ 2G^{AB} \frac{V_{,AB}}{V} - \frac{5}{2} G^{AB} \frac{V_{,A}V_{,B}}{V^2} \right\} \quad (19)$$

The Lyapunov exponent is then given by

$$\lambda = \lim_{t \rightarrow \infty} \sqrt{-\tilde{R}/3} < \sqrt{T} > \quad (20)$$

where $< T >$ is a mean kinetic energy. This expression is then used by Szydlowski and Lapeta to give an invariant characterization of chaos in Bianchi IX. The problem is however that this procedure breaks down for the Bianchi IX model. To see this, recall that the Maupertius principle is applicable in situations where the potential V is either positive or negative definite. This is not the case for the Bianchi IX model; hence the necessity of taking the modulus of V in equation (18). In changing sign, V must pass through zero at which point the conformal transformation (18) breaks down; i.e. all distances in one space become compacted to zero length in the conformally related manifold. The Bianchi IX potential is not positive definite and neither can it be made so by a re-scaling of the definition of V . This is because in using the Hamiltonian formulation we have had to impose constraints which stop us making a suitable re-scaling. Since V takes on both positive and negative values, it also passes through zero and at such points \tilde{R} becomes infinite (see also Figure 2 in Szydlowski and Lapeta 1990). Furthermore, the $V = 0$ equipotential is an open curve and so all trajectories will cross this curve an infinite number of times as the singularity is approached. Burd and Tavakol numerically evolve the equations of motion and show that this is indeed the case.

4. OTHER AVENUES

One frequently wishes to analyse systems which are perturbations of an integrable Hamiltonian. In this case, a suite of techniques has been developed which allow one to

ask whether such a system is chaotic, and for what range of parameter values it is chaotic (see e.g. Lichtenberg and Lieberman 1983, Guckenheimer and Holmes 1983, Wiggins 1990 and the contribution by Calzetta in this volume). These techniques rely partly upon the geometrical nature of the trajectories in the phase space and in particular upon the existence (or otherwise) of a homoclinic trajectory and the evaluation of the Melnikov integral. Such techniques should be gauge invariant and have already been successfully employed in examining chaotic particle trajectories around black holes (Bombelli and Calzetta 1992) and for the study of scalar fields in Friedmann-Robertson-Walker cosmologies (Calzetta and Hasi 1993).

So far these methods have not been applied directly to the Bianchi IX cosmology though work by Koiller *et al* (1985) may be relevant here. These authors consider spacetimes with a spatial metric which can be split to form a product space, $S^2 \times S^1$. The integrable case is taken to be a static model and perturbations of the S^2 and S^1 spaces are considered both separately and together. When both perturbations are switched on the authors show the existence of a Smale Horseshoe which heralds the onset of chaos.

Another possible approach to deciding on whether or not the Bianchi IX models are chaotic is to examine the integrability of the dynamical system itself. One technique for doing this is to study the analytic structure of the dynamical system in the complex time plane by making use of Painlevé singularities (see e.g Ramani *et al* 1989). This approach has been used by Cotsakis (Cotsakis 1990) to study the integrability of the Euclidean Bianchi IX cosmology. Using this technique he was able to show that this model was integrable. One should be aware, however, that the use of Painlevé singularities has its attendant pitfalls; if it shows that the system is indeed integrable, then this will be the case, but if it shows the opposite, the system need not be chaotic (see the article by Ramani for more details of this).

5. CONCLUSIONS

To strictly show that a dynamical system is chaotic requires one to demonstrate that there is a sensitive dependence upon initial conditions (for example that at least one of the Lyapunov exponents is positive) and that the phase space is compact. In this strict sense, there has yet to be a rigorous demonstration of the chaotic nature of the Bianchi IX model. Those approaches, such as Chitre's, which use a compact phase space are also generally asymptotic approximations to the full flow. One interesting thing to emerge from numerical calculations of the Bianchi models is just how well the BKL map approximates the flow. As the singularity is approached one finds that the evolution follows that predicted from the map to a very high accuracy (Berger 1991). Of course far away from the singularity the assumptions underlying the map break down (Berger 1990, 1991 and Hobill *et al* 1991). It may be that this is the best that we can do.

Whilst the Lyapunov exponent gives a quantitative measure of the chaotic behaviour, the study of homoclinic trajectories seems to provide a gauge invariant though qualitative description of the chaos. One line of work being pursued at the moment is to use the Melnikov method to examine what happens when one represents the Bianchi IX mode as a closed FRW model with gravitational waves on it.

This study has shown that one must think quite carefully about what it really means to describe a system as chaotic. Usually it means that neighbouring trajectories

diverge exponentially from each other so that systems will appear to rapidly lose the memory of their initial conditions. What is important is the relative timescales involved. This can also provide a motivation for further research into this area. Whilst the formulation of a robust, gauge invariant measure of chaos is important, another avenue is to ask what are the physical consequences of the behaviour shown by this model. It has been known for a long time that there can have been only a few bounces between the end of the quantum era and now (Doroshkevich and Novikov 1971). But one can ask what would happen to say the microwave background if such a bounce were to have occurred between the time of last scattering and now; what would be the effect if a bounce occurred during the era of nucleosynthesis.

ACKNOWLEDGEMENTS

I would like to thank my collaborators Reza Tavakol, Nic Buric and George Ellis as well as David Hobill, Beverly Berger, Sven Rugh and Jim Skea for many interesting and enlightening discussions.

REFERENCES

- Arnold, V., I., 1989, *Mathematical Methods of Classical Mechanics* 2nd ed (Springer Verlag).
- Barrow, J., D., 1982, *Phys. Rep.*, **85**, 1.
- Belinski, V., A., Khalatnikov, I., M. and Lifshitz, E., 1970, *Adv. Phys.*, **19**, 525.
- Berger, B., K., 1990, *Class. Quantum Grav.*, **7**, 203.
- , 1991, *Gen. Rel. Grav.*, **23**, 1385.
- Bombelli, L., and Calzetta, E., 1992, *Class. Quantum Grav.*, **9**, 2573.
- Burd, A., B., Buric, N. and Ellis, G., F., R., 1990, *Gen. Rel. Grav.*, **22**, 349.
- Burd, A., B., Buric, N. and Tavakol, R., K., 1991, *Class. Quantum Grav.*, **8**, 123.
- Calzetta, A and El Hasi, C., 1983, *Chaotic Friedmann- Robertson-Walker Cosmology* (preprint).
- Chitre, D., M., 1972, *Investigation of Vanishing of a Horizon for Bianchi Type IX* (University of Maryland, Dpt. of Physics Technical Report No. 72-125).
- Cotsakis, S., 1990, *Cosmological Models in Quantum Gravity*, (University of Sussex, D. Phil. Thesis)
- Doroshkevich, A. G. and Novikov, I. P., 1971, *Sov. Astron.*, **14**, 763.
- Ellis, G., F., R. and MacCallum, M., A., H., 1969, *Commun. Math. Phys.*, **12**, 108.
- Francisco, G. and Matsas, G., E., A., 1988, *Gen. Rel. Grav.*, **20**, 1047.
- Guckenheimer, J. and Holmes, P., 1983, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, (Springer Verlag)
- Hedlund, G., A., 1939, *Bull. Am. Math. Soc.*, **45**, 241.
- Hobill, D., Bernstein, D., Welge, M. and Simkins, D., 1991, *Class. Quantum Grav.*, **8**, 1155.
- Holmes, P., 1990, *Phys. Rep.*, **193**, 137.
- Hopf, E., 1936, *Trans. Am. Math. Soc.*, **39**, 299.

- Khalatnikov, I., M., Lifshitz, E., M., Khanin, K., M., Schur, L., N. and Sinai, Ya, G., 985, *J. Stat. Phys.*, **38**, 97.
- Koiller, J., De Mello Neto, J., R., T. and Damião Soares, 1985, *Phys. Lett.* , **110A**, 260.
- Lichtenberg, A., J. and Lieberman, M., A., 1983, *Regular and Stochastic Motion*, (Springer Verlag)
- Mayer, D. H., 1987, *Phys. Lett.*, **121A**, 390.
- Oseledec, V. I., 1968, *Trans. Moscow Math. Soc.*, **19**, 197.
- Palmer, T., N., 1993, *Bull. Am. Meteorological Soc.*, **74**, 49.
- Pesin, Ya. B., 1977, *Usp. Mat. Nauk.*, **32**, 55.
- Poincaré, H., 1899, *Les Méthodes Nouvelles de la Mécanique Céleste*, vols 1–3, (Gauthier Villare, Paris)
- Pullin, J., 1990, *Syracuse Report No. 90-0734*, (unpublished)
- Ramani, A., Grammaticos, B. and Bountis, T., 1989, *Phys. Rep.*, **180**, 159.
- Rugh, S., *Chaotic Behaviour and Oscillating Three Volumes in a Space Time Metric in General Relativity*, (Cand. Sci. Thesis, Niels Bohr Institute).
- Rugh, S., E. and Jones, B., J., T., 1990, *Phys. Lett.* , **147A**, 353.
- Szydlowski, M. and Lapeta, A., 1990, *Phys. Lett.* , **148A**, 239.
- Wainwright, J., 1987, *Proceedings of the Second Hungarian Relativity Workshop*, ed. Z. Perjés, (World Scientific).
- Wiggins, S., 1990, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, (Springer Verlag).
- Zardecki, A., 1983, *Phys. Rev.* , **D28**, 1235.

THE CHAOTICITY OF THE BIANCHI IX COSMOLOGICAL MODEL

Gerson Francisco¹ and George E.A. Matsas²

Instituto de Física Teórica-UNESP
Rua Pamplona 145, 01405-900-São Paulo, São Paulo, Brazil

Abstract. We briefly review the primeval role played by the Bianchi IX model in understanding the chaotic nature of the Einstein equations.

The last years have witnessed an increasing interest in investigating the chaotic nature of the Einstein equations. The organization of a workshop dedicated to this particular issue, and the growing number of research papers in this area are clear indications of the necessity of introducing a suitable concept of chaos in General Relativity. As far as we know, the Bianchi IX model was the prime solution of Einstein equations which had its chaotic properties studied. Actually, this model has played an outstanding role in the investigation of chaos in the context of General Relativity. In this vein, a brief review on the development of our understanding concerning the chaotic (and non-chaotic) features of the Bianchi IX model may be in order. This paper is an incomplete tentative to accomplish this goal.

The Bianchi IX model was a source of intense investigation over the sixties by the Russian school in trying to understand the nature of the cosmological singularities which arise from Einstein equations (see Belinkii et al [1] and references therein), and by Misner in trying to solve the cosmological horizon problem [2]. As a result, it was realized that there were stochastic discrete maps associated with this model [3] (see also Barrow [4] for a comprehensive discussion). Close to the singularity, gravity should dominate over matter, and so we are lead to restrict our analysis to the vacuum case.

The Bianchi IX model is a homogeneous but non-isotropic solution of Einstein equations based in the $SO(3)$ isometry group. The diagonal Bianchi IX model, i.e. Mixmaster universe, is characterized by its spatial metric components (a^2, b^2, c^2) . As one approaches the singularity, the Bianchi IX model behaves as following a succession of *Kasner epochs*. The spacetime associated with the vacuum Kasner universe is

¹e-mail: GF_Akham@IFT.UESP.ANSP.BR

²e-mail: Matsas@IFT.UESP.ANSP.BR

characterized by just one parameter $u \in \mathbf{R}_+$. Hence, the Bianchi IX model can be approximately described close to the singularity by giving the set of Kasner parameters $\{u_n\}$ associated with it. The map which rules the transition from one *Kasner epoch* to the next one is

$$u_{n+1} = u_n - 1 \quad \text{if } 2 \leq u_n < +\infty, \quad (1)$$

and

$$u_{n+1} = \frac{1}{u_n - 1} \quad \text{if } 1 < u_n \leq 2. \quad (2)$$

Beginning with an arbitrary $u_0 > 2$, a number of $[u_0]$ (i.e., integer part of u_0) different *Kasner epochs* occur before transition (2) takes place, and a new *Kasner era* starts again. We will assume that u_0 is irrational, otherwise the oscillatory pattern of *Kasner eras* would stop after a finite number of transitions. This is also natural on mathematical grounds since the set of rational numbers has null measure in the set of real numbers. From (1) and (2) it can be shown that the number of *Kasner epochs* present in the *Kasner era* $m + 1$ is $[1/x_m]$, where $x_0 = u_0 - [u_0]$, and

$$x_{m+1} \equiv T(x_m) = x_m^{-1} - [x_m^{-1}]. \quad (3)$$

This discrete map $T : [0, 1] \rightarrow [0, 1]$, with $T(0) \equiv 0$ was analytically studied, and its Liapunov exponent (LE) calculated

$$\lambda \equiv \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N \log |T'(x_n)| = \frac{\pi^2}{6(\ln 2)^2}. \quad (4)$$

Since λ is positive, it was concluded that the information on the initial conditions is exponentially erased. In the early eighties, the Bianchi IX model was already considered a prototype chaotic solution of Einstein equations, and other discrete maps associated with this model were studied [5].

The first one to perform a numerical analysis of the principal LE as computed directly from the differential equations which govern the Bianchi IX dynamics was Zardecki [6]. He was mostly concerned with the dependence of the chaotic behavior of the Bianchi IX model with respect to introducing random-noise terms. His numerical results seemed to confirm the well established belief that the principal LE obtained by time evolving these equations should be non-zero.

Notwithstanding, some years later when the present authors were studying the Bianchi IX behavior close to the singularity using qualitative methods of dynamical systems due to Bogoyavlenskii and Novikov [7], it was noticed that the LE obtained by evolving the differential equations in the Ω -time should not converge to a stable fixed positive value. (The relation between the intrinsic time Ω , and the comoving proper time t is given by $d\Omega = -(abc)^{-1}dt$.) The explanation of this phenomenon, as discussed in [8], was based on the fact that the Ω -time spent by the universe in some *Kasner era* n increases as

$$\Omega_n \sim e^{\alpha n}, \quad (5)$$

where $\alpha \sim 2$ (see Ref. [9]). As a result, successive transitions ruled by map (3) take longer and longer to occur in Ω -time, and the maximal LE for the continuous orbit must necessarily go to zero.

At the time this result was being worked out, the main computer facility available was a PC-XT clone. A rough numerical analysis was performed using Kasner initial conditions, yet not satisfying the constraint associated with $R_0^0 = 0$. It was enough to

corroborate the expectation that the LE calculated with respect to the Ω -time should decrease, in accordance with general arguments based on (5).

One year later the anomalous behavior of the LE was confirmed by a number of researchers in the 12th International Conference on General Relativity held at Boulder, Colorado, and detailed papers were soon published [10, 11] (see also Ma and Wainwright [12]). Distinct aspects of the chaotic behavior of the Bianchi IX model have been clarified since then. The LEs obtained by time evolving the differential equations with initial conditions satisfying the constraint (modulo machine precision) were shown to tend to zero quickly. Actually, it was argued [13, 11] that Kasner initial conditions, although natural under the point of view of the previous discussion, are not suitable to be used in the calculation of the LEs. This is so because they induce the appearance of an effective negative energy density which produces an unphysical enhancement of the Bianchi IX chaoticity. A different approach was used by Berger who studied the chaoticity of the Bianchi IX model regarded as evolving in the minisuperspace [14]. Also, efforts were spent to investigate the relationship between the dynamics of the maps employed to approximate the flow and the dynamics of the flow itself [15].

At this point, it was already clear that the problem in bringing the concept of chaos to General Relativity was related with the general covariance of the theory: The LE concept is strongly dependent on the time parameter chosen to evolve the system [16]. This problem does not seem likely to be solved by just favoring one particular time parameter (say, proper time) not only because it goes against the whole philosophy underlying General Relativity, but also because the concept of time close to the cosmological singularity becomes misty (see Ref. [17], p.813). Short time averages lead, in principle, to a better understanding of the chaos present in non-autonomous systems than long time averages [10]. A proposal for classifying the LE behavior for non-autonomous systems has been presented in [18]. Some attempts in treating the problem in terms of invariants, e.g. scalar curvature, etc, are in progress (see Szydłowski [19] and references therein, but see also Burd and Tavakol [20]).

Summarizing, the Bianchi IX model has associated with it some discrete maps which can be properly studied, and are clearly chaotic. However, it seems to exist, up to the present, no satisfactory general procedure to study the chaotic behavior of the cosmological solutions of Einstein equations. The best one can do is to investigate different aspects of each model concerning its chaoticity, rather than labeling the model as chaotic or non-chaotic as a whole. Undoubtedly, it seems fair to say that General Relativity is expected to play a prominent role in the forthcoming development of the chaos concept.

ACKNOWLEDGEMENTS

We are indebted to the organizer committee of the NATO Advanced Research Workshop - Deterministic Chaos in General Relativity, and to David Hobill, in particular, who invited us to write this paper. We acknowledge Coordenadoria de Desenvolvimento Científico e Tecnológico for partial (GF), and full financial support (GM).

REFERENCES

- [1] Belinskii, V.A., Lifshitz E.M., and Khalatnikov I.M., 1970, *Usp. Fiz. Nauk.*, **102**, 463 [1971, *Sov. Phys. Usp.*, **13**, 745]. Belinskii, V.A., Khalatnikov,I.M., and Lifshitz E.M., 1970, *Adv. Phys.*, **19**, 525.

- [2] Misner, C.W., 1969, *Phys. Rev. Lett.*, **22**, 1071.
- [3] Lifshitz, E.M., Lifshitz, I.M., and Khalatnikov, I.M., 1970, *Zh. Eksp. Teor. Fiz.*, **59**, 322 [1971, *Sov. Phys. JETP*, **32**, 173].
- [4] Barrow, J.D., 1982, *Phys. Rep.*, **85C**, 1.
- [5] Chernoff, D.F., and Barrow, J.D., 1983, *Phys. Rev. Lett.*, **50**, 134.
- [6] Zardecki, A., 1983, *Phys. Rev. D*, **28**, 1235.
- [7] Bogoyavlenskii, O.I., 1985, *Qualitative Theory of Dynamical Systems in Astrophysics and Gas Dynamics*, (New York: Springer Verlag).
- [8] Francisco, G., and Matsas, G.E.A., 1988, *Gen. Rel. Grav.*, **20**, 1047.
- [9] Khalatnikov, I.M., Lifshitz, E.M., Khanin, K.M., Schur, L.N., and Sinai, Ya.G., 1985, *J. Stat. Phys.*, **38**, 97.
- [10] Burd, A.B., Buric, N., and Ellis, G.F.R., 1989 *Abstract of 12th Intern. Conf. on Gen. Relat.* held at Boulder (CO); Burd, A.B., Buric, N., and Ellis, G.F.R., 1990, *Gen. Rel. Grav.*, **22**, 349.
- [11] Hobill, D., Bernstein, D., Welge, M., and Simkins, D., 1989, *Abstract of 12th Intern. Conf. on Gen. Relat.* held at Boulder (CO); Hobill, D., Bernstein, D., Welge, M., and Simkins, D., 1991, *Class. Quant. Grav.*, **8**, 1155.
- [12] Ma, P.K-H., and Wainwright, J., 1989, *Proceed. of the Third Hungarian Relativity Workshop, Budapest*, (unpublished).
- [13] Rugh, S.E., and Jones, B.J.Y., 1990, *Phys. Lett. A*, **147**, 353.
- [14] Berger, B.K., 1990, *Class. Quant. Grav.* **7**, 203.
- [15] Burd, A.B., Buric, N., and Tavakol, R.K., 1991, *Class. Quant. Grav.*, **8**, 123. Berger, B.K., 1993, *Phys. Rev. D*, **47**, 3222.
- [16] Pullin, J., 1991, in *SILARG VII Relativity and Gravitation: Classical and Quantum*, edited by J.C. D'Olivio, E. Nahmad-Achar, M. Rosenbaum, M.P. Ryan,Jr., L.F. Urrutia, and F. Zertuche (World Scientific, Singapore); Ferraz, K., Francisco, G., Matsas, G.E.A., 1991, *Phys. Lett. A*, **156**, 407. Berger, B.K., 1991, *Gen. Rel. Grav.*, **23**, 1385.
- [17] Misner, C.W., Thorne, K.S., Wheeler, J.A., 1973, *Gravitation* (Freeman, San Francisco).
- [18] Ferraz, K., and Francisco, G., 1992, *Phys. Rev. D*, **45**, 1158.
- [19] Szydlowski, M., 1993, *Phys. Lett. A* **176**, 22.
- [20] Burd, A., and Tavakol, R., 1993, *Phys. Rev. D*, **47**, 5336.

CHAOS IN THE EINSTEIN EQUATIONS - CHARACTERIZATION AND IMPORTANCE?

Svend E. Rugh¹

The Niels Bohr Institute, University of Copenhagen,
Blegdamsvej 17, 2100 København Ø, Denmark

Abstract. Is it possible to define what we could mean by chaos in a space-time metric (even in the simplest toy-model studies)? Is it of importance for phenomena we may search for in Nature?

1. INTRODUCTION

Theoretical physics would not die out even if we had *already* found the “master plan”, the “master law” (T.O.E.) for how the Universe in which we live is constructed. Namely, given the knowledge of the basic laws of physics (on each level on the Quantum Staircase, say) it is still a major project to try to deduce their *complex* consequences, i.e. to find out the complex ways in which matter (and their interactions) organizes itself into living and nonliving forms.

The word “chaos” is somewhat of an unlucky choice since we may too easily associate it with something which is structureless. Chaotic systems do not lack structure. On the contrary, chaotic systems exhibit far more interesting and richer structure in their dynamical behavior than integrable systems.

However, chaotic systems have an aspect of unpredictability and “simulated” disappearance of information: Information walks down to the small scales and is replaced by noise walking up from the small scales. I.e. chaos pumps *information* up and down the chain of decimals in the phase space coordinates! (due to the exponential amplification of uncertainties in the specification of the initial state).

Many physical systems, governed by *non-linear* equations of motion, will exhibit chaos. By now, examples are known from all disciplines in physics.

¹e-mail: rugh@nbivax.nbi.dk

Due to the highly non-linear *self-interaction* of the gravitational and non-Abelian gauge fields, the time evolution of (generic) configurations of “gauge fields” and “gravitational fields” will be non-integrable even without any coupling to material bodies. For example, as evidenced by the simplest toy-model studies, we expect that the Einstein equations - in scenarios involving *strong* field strength, where the non-linearity of the equations is important (e.g. probing the Einstein equations near space-time singularities) - will exhibit *chaotic solutions* (rather than integrable ones) if not too much symmetry is imposed on the field configuration. The same is true for non-Abelian gauge fields. While on one hand this is *not very surprising* (the Einstein and Yang-Mills equations are highly *non-linear* theories and one of the lessons from chaos theory has been that *even the simplest* non-linear equations usually exhibit “chaos”!) there are, on the other hand, several issues of interest:

1. Can “standard indicators of chaos” be used to characterize the “metrical chaos”? In other words: Are there some *deeper problems* in characterization of chaos in this context compared to other chaotic physical models?
2. What is the *structure* of this (non-dissipative) chaos? How does chaos “look like” in simple toy-models of the classical Yang-Mills equations and the classical Einstein equations?
3. Is it of any *physical significance* that field configurations of the fundamental forces at the classical (or semi-classical) level may exhibit chaotic, irregular non-integrable solutions?

Let us turn our attention to the gravitational field: It is well known that individual orbits of test particles (bodies) in a *given* gravitational field can exhibit “chaos”. This is the case for the motion of test particles in Newton’s theory of gravitation (e.g. “chaotic” orbits of individual stars in the potential generated by the other stars in a galaxy) and also in general relativity: Cf., e.g., the study by G. Contopoulos of periodic orbits and chaos around two black holes, G. Contopoulos (1990), the study of chaos around a single black hole, L. Bombelli and E. Calzetta (1992) and the study of chaotic motion of test particles around a black hole immersed in a magnetic field, V. Karas *et al* (1992). Also an extended object like a (cosmic) string may jump chaotically around the equatorial plane of a black hole, cf. A.L. Larsen (1993).

Because of their nonlinearity, the Einstein field equations however permit space-time to be curved (“gravity generates gravity”) - even in the absence of any nongravitational energy - and the dynamical evolution of the spacetime metric “ $g_{\mu\nu}(x)$ ” itself is governed by *highly nonlinear* equations and may allow solutions of the chaotic type for the “metric” field itself.

If we explicitly write down (cf. Kip S. Thorne (1985)) the vacuum Einstein equations $G_{\mu\nu} = 0$ as differential equations for the “metric density” $\tilde{g}^{\mu\nu} \equiv \sqrt{-g} g^{\mu\nu}$ we have a very complicated set of partial differential equations for $\tilde{g}^{\mu\nu}$,²

$$\begin{aligned}\tilde{g}^{\mu\nu} \frac{\partial^2 \tilde{g}^{\alpha\beta}}{\partial x^\mu \partial x^\nu} &= \tilde{g}_{,\mu}^{\alpha\nu} \tilde{g}_{,\nu}^{\beta\mu} + \frac{1}{2} \tilde{g}^{\alpha\beta} \tilde{g}_{\lambda\mu} \tilde{g}_{,\rho}^{\lambda\nu} \tilde{g}^{\rho\mu} \\ &+ \tilde{g}_{\lambda\mu} \tilde{g}_{,\nu}^{\nu\rho} \tilde{g}_{,\rho}^{\alpha\lambda} \tilde{g}^{\beta\mu} - \tilde{g}^{\alpha\lambda} \tilde{g}_{\mu\nu} \tilde{g}_{,\rho}^{\beta\nu} \tilde{g}^{\mu\rho} - \tilde{g}^{\beta\lambda} \tilde{g}_{\mu\nu} \tilde{g}_{,\rho}^{\alpha\nu} \tilde{g}_{,\lambda}^{\mu\rho}\end{aligned}$$

²Note, that $\tilde{g}_{\mu\nu}$ (the inverse of $\tilde{g}^{\mu\nu}$) is a highly nonlinear algebraic function of $\tilde{g}^{\mu\nu}$. Repeated indices are to be summed, commas denote partial derivatives, e.g. $\tilde{g}_{,\mu}^{\alpha\beta} \equiv \partial \tilde{g}^{\alpha\beta} / \partial x^\mu$. The coordinate system has been specialized so the metric is in “deDonder gauge”, $\tilde{g}_{,\beta}^{\alpha\beta} = 0$ (see K.S. Thorne (1985)).

$$+ \frac{1}{8}(2\tilde{g}^{\alpha\lambda}\tilde{g}^{\beta\mu} - \tilde{g}^{\alpha\beta}\tilde{g}^{\lambda\mu})(2\tilde{g}_{\nu\rho}\tilde{g}_{\sigma\tau} - \tilde{g}_{\rho\sigma}\tilde{g}_{\nu\tau})\tilde{g}^{\nu\tau}_{,\lambda}\tilde{g}^{\rho\sigma}_{,\mu} . \quad (1)$$

The left hand side of (1) is a kind of curved spacetime wave operator “ \square ” acting on $\tilde{g}^{\alpha\beta}$ (giving a propagation effect of the gravitational degrees of freedom) whereas the right hand side is a sort of “stress-energy pseudotensor” for the gravitational field which is quadratic in the first derivatives of $\tilde{g}^{\alpha\beta}$ and acts as the source for “ \square ” $\tilde{g}^{\alpha\beta}$.

At first, general relativity, i.e. Einsteins theory of gravitation, is not even a dynamical theory in the usual sense. It does not, from the very beginning, provide us with a set of parameters (describing the gravitational degrees of freedom) evolving in “time”. “Time” loses here its absolute meaning as opposed to the classical dynamical theories where the “Newtonian time” is taken for granted. The division between space and time in general relativity comes through foliating the space-time manifold \mathcal{M} into spacelike hypersurfaces Σ_t . The metric $g_{\mu\nu}$ on \mathcal{M} induces a metric g_{ij} on Σ_t and can be parametrized in the form

$$g_{\mu\nu} = \begin{pmatrix} N_i N^i - N^2 & N_j \\ N_i & g_{ij} \end{pmatrix}$$

bringing the metric on the 3 + 1 form³

$$ds^2 = -N^2 dt^2 + g_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \quad (2)$$

where N and N_i are called lapse function and shift vector respectively. (Cf, e.g., MTW §21.4)

Only after splitting the space-time into space and time (the 3+1 ADM splitting) we yield the possibility to treat the evolution of the metric under the governing Einstein equations as a dynamical system on somewhat equal footing as other dynamical systems which have some (physical) degrees of freedom evolving in “time”.⁴

It is of interest to our discussion on the dynamics of the Einstein equations (chaotic or not) to know that one may show that the Einstein equations indeed admit a well posed *initial value* formulation, so the Einstein equations do *determine* the evolution of the metric (up to gauge transformations) *uniquely* from given initial conditions and the solutions admit a *Cauchy stability criteria* which establishes that the solutions depend *continuously* on the initial data. Cf., e.g., discussion in MTW §21.9, S.W. Hawking and G.F.R. Ellis (1973), sec.7, and R.M. Wald (1984), sec.10.

Given the complexity of the Einstein equations (1) by comparison, e.g., with the Navier-Stokes equations, it is not surprising that we at present have only little understanding of the dynamics of the Einstein equations involving strong gravitational fields.

³Albert Einstein taught us to treat space and time on an equal footing in a four-dimensional spacetime manifold and build up a space-time covariant formulation of the theory, yet to deal with dynamics we manifestly have to break the space-time covariance of the formulation.

⁴Compare, e.g., with the characterization of turbulence in connection with fields like Navier-Stokes flows or $SU(2)$ Yang Mills fields. In these cases there are no such fundamental problems or ambiguities as concerns the fields being in certain well defined space points and evolving in a well defined “external” time t . Such fields evolve in a flat Minkowskij metric (if they are not coupled to gravitational fields) and no “mixing” between space and time concepts (up to “stiff” Lorentz transformations) occurs.

From the observational side not many observable phenomena which involve the gravitational field need the full non-linear Einstein equations and in “daily life” gravity often the *linearized* equations suffice.⁵ Among the three *classic* tests of Einstein’s theory only the precession of the perihelia of the orbits of the inner planets tests a non-linear aspect of the Einstein equations.

To find scenarios where the full non-linearities of the Einstein equations are important one has to search among astrophysical and cosmological phenomena far removed from “daily life” gravity.

Among non-linear phenomena in general relativity (geons, white and black holes, wormholes, cf. Kip S. Thorne (1985), solitons, cf. G.W. Gibbons (1985)) the formation of spacetime singularities is one of the most remarkable phenomena which appears in nonlinear solutions of the *classical* Einstein equations under a variety of circumstances, cf. the singularity theorems by S.W. Hawking and R. Penrose (1970).

The singularity theorems of Hawking and Penrose tell us, however, practical nothing about the structure of such spacetime singularities. What do they look like?

1.1. Highly symmetric gravitational collapses as a laboratory for testing ideas about how to characterize “chaos” in a general relativistic context.

To capture a nonlinear aspect of the Einstein equations it is natural to probe them in scenarios involving strong field strength, e.g. near curvature singularities.

Without some restrictions of symmetry imposed on the spacetime metric $g_{\mu\nu}(x)$ the Einstein equations are intractable (though considerable progress has been made in numerical relativity of solving cases with little symmetry).

We shall consider a simple example and use it to test ideas and concepts about chaos. (If we are not able to agree upon how chaos should be defined in this simple example we may very well give up all hope to develop indicators of chaotic behavior in more complicated examples of spacetime metrics).

Thus, we restrict, for simplicity, attention to toy-model metrics with *spatially homogeneous* three-dimensional space-like slices (hypersurfaces): Then the gravitational fields are the same at every point on each of the surfaces of homogeneity and one may thus represent these fields via *functions of time* only!

More explicitly, spatially homogeneous 3-geometries are 3-manifolds on which a three-dimensional Lie group acts transitively. On the 3-manifold this symmetry is encoded in the existence of three linearly independent spacelike Killing vectors ξ_i , $i = 1, 2, 3$, satisfying the Lie algebra $[\xi_i, \xi_j] = C_{ij}^k \xi_k$ where C_{ij}^k are the structure constants of the Lie algebra.⁶

The particular collapse (“big crunch”), the mixmaster collapse, we shall consider has the same non-Abelian isometry group $SU(2)$ on the three-space (i.e. same topology $\sim R \times S^3$) as the compact FRW-collapse but it contains three scale-factors a, b, c instead of just one. A ‘*freely falling astronomer*’ who falls into the spacetime singularity of the

⁵Note, however, that *nonlinear effects* of the Einstein equations may show up to be important even in regions where one would imagine the linearized equations to be sufficient. For example, one has a *nonlinear* effect in the form of a permanent displacement of test masses (of the same order of magnitude as the *linear* effects) after the passage of a gravitational wave train - even when the test masses are placed at arbitrary distances from the gravitational wave source. See D. Christodoulou (1991).

⁶The classification of three dimensional Lie-algebras dates back to L. Bianchi (1897) and the spatially homogeneous metrics are therefore often referred to as Bianchi metrics. See e.g. M.A.H. MacCallum (1979, 1983).

'big crunch' will experience a growing tidal field, in which he is *compressed* along two directions and *stretched* (expanded) in one direction, the directions being permuted infinitely many times in a *not-predictable* way.

The possibility of chaos has been investigated only for very few toy model studies of spacetime metrics! Whether one should expect the Einstein equations to generate "chaos" in generic cases (for strong gravitational fields, i.e., high curvatures) the answer is absolutely: Nobody knows!

The paper is organized as follows: In sec.2 we describe various aspects of the mixmaster gravitational collapse. Not surprisingly, a collapse to a spacetime singularity is prevented if one includes matter with negative energy and pressure. In that case, however, the behavior of the spacetime metric is very interesting, very irregular and highly unpredictable, and oscillations of the three-volume occur. (Due to the negative pressure and energy density the attraction of matter turns into an unphysical repulsion preventing the "universe" from collapsing). If the metric is evolved according to the vacuum Einstein equations, the dynamics has, after some transient, a monotonically declining three-volume and the degrees of freedom of the spacetime metric is fast attracted into an interesting self-similar, never ending oscillatory behavior on approach to the big crunch singularity. This may be understood as a never ending sequence of short bounces against a potential boundary generated by the three-curvature scalar (^3R) on the three-space. This scattering potential becomes, to a very good approximation, infinitely hard when the metric approaches the singularity, and the collapse dynamics may, in that limit, be captured by a set of simple algebraic transition rules (maps), for example the so-called "Farey map", which is a strongly intermittent map (this map has, as sub-map, the Gauss map which is well known in chaos theory and which has positive Kolmogorov entropy).⁷

In sec.3 we describe the problem - inherent to general relativity - of transferring standard indicators of chaos, in particular the spectrum of Lyapunov exponents, to the general relativistic context, since they are highly gauge dependent objects. This fact was pointed to and emphasized in S.E. Rugh (1990 a,b). I can only moderately agree that this observation was arrived at independently by J. Pullin (1990). By referring to a specific gauge (the Poincaré disc) Pullin misses the point (in my opinion). No "gauge" is better than others. One should try to develop indicators which capture chaotic properties of the gravitational field ("metric chaos") in a way which is invariant under spacetime diffeomorphisms - or prove that this can not be done! (H.B. Nielsen and S.E. Rugh). This program of research is still in its infancy.

In sec.4 and sec.5 some (even) more wild speculations are offered concerning the generality and applicability of the concept of "metrical chaos" etc. One would like to argue - but it is not easy - that *non-integrability* of the Einstein equations, is a generic phenomenon when considering scenarios involving really strong gravitational fields, e.g. near Planck scales where the gravitational field should be treated quantum mechanically. Whether there are implications of "metrical chaos" on the quantum level is a question which not possible to address since no good candidate for a theory of quantum gravity is known. In the context of the Wheeler-DeWitt equation (which however involves arbitrariness, cf. e.g. the factor ordering problem) one may address this question for the mixmaster gravitational collapse. This is beautifully illustrated in the "Poincaré disc gauge" (which I describe shortly) and was already considered by Charles W. Misner

⁷Results obtained in the first part of sec.2 were also arrived at by D. Hobill *et al* in completely independent investigations.

twenty years ago. The mixmaster collapse dynamics is however so special (the scattering domain of the Poincaré disc tiles the disc under the action of an “arithmetic group”) that its quantization exhibits ungeneric features, relative to more generic Hamiltonian models (of similar low degree of dimensionality) studied in the discipline of “Quantum Chaos”⁸. This illustrates, once again, that the mixmaster gravitational collapse is a very beautiful, yet very special, example of chaos (algebraic chaos). However, for our purpose, to use it as a toy-model to investigate the applicability of indicators of chaos in the general relativistic context, it serves as a good starting point.

It is interesting whether “metrical chaos” (not yet defined) has potential applications for phenomena occurring in Nature. Certainly, non-integrability (i.e. lack of first integrals relative to the number of degrees of freedom) and non-linear effects may show up even in scenarios involving rather weak fields (cf. e.g. D. Christodoulou (1991)).

Considering the possibility of the early Universe to be described by the mixmaster metric, we note in sec.6 that, according to the Weyl curvature hypothesis (of R. Penrose), which suggests that the Weyl curvature tensor should vanish at the initial singularity (at “big bang”), the mixmaster metric has too big Weyl curvature to be implemented in our actual Universe at Planck scales, say. The Guth/Linde inflationary phase may modify this viewpoint.

In sec.6 also some more general reflections are put forward concerning the “chaotic cosmology” concept by Charles W. Misner *et al* which attempt at arriving at our present Universe from (almost) arbitrary initial conditions.

2. THE MIXMASTER GRAVITATIONAL COLLAPSE GIVES A “HINT” OF THE SORT OF COMPLEXITY (“METRICAL CHAOS”) ONE MAY HAVE FOR SOLUTIONS TO EINSTEIN EQUATIONS.

The mixmaster gravitational collapse is a very famous⁹ gravitational collapse (a “big crunch”) which generalizes the compact FRW collapse and which gives us a “hint” of the sort of complexity (“metrical chaos”) one should expect for gravitational collapses which have more degrees of freedom than the simple (integrable) FRW-collapse.

We imagine that a “3+1” split has been performed, splitting the spacetime manifold into the topological product of a line (the “time” axis) and the three-dimensional spacelike hypersurfaces Σ_t (the dynamical degrees of freedom are the *spatial* components of the metric, the induced metric g_{ij} on Σ_t , which evolves in the “time” parameter “ t ”). In fact, we shall operate in a “synchronous reference frame” (Landau and Lifshitz §97 and MTW, §27.4) which brings the spacetime metric (2) on the very simple form $ds^2 = -dt^2 + g_{ij}dx^i dx^j$.

⁸“Quantum Chaos”, or what Michael Berry has named “Quantum Chaology”, cf. e.g. M. Berry (1987), investigates the semi-classical or quantum behavior characteristic of Hamiltonian systems whose classical motion exhibits chaos.

⁹Cf., e.g., Ya.B. Zel’dovich and I.D. Novikov (1983), esp. §22; C.W. Misner, K.S. Thorne and J.A. Wheeler (1973), esp. §30 or L.D. Landau and E.M. Lifshitz (1975), esp. §116-119. See, also, J.D. Barrow (1982).

2.1. What are the degrees of freedom in this toy model?

The symmetry-ansatz for the metric¹⁰ is:

$$ds^2 = -dt^2 + \gamma_{ij}(t)\omega^i(x)\omega^j(x) , \quad \gamma_{ij}(t) = \text{diag}(a^2(t), b^2(t), c^2(t)) \quad (3)$$

The spacetime has the topology $\mathbf{R} \times S^3$ (product of a time axis and the compact three-sphere). The three-space is invariant under the $G=\text{SU}(2)$ group, as expressed by the $\text{SU}(2)$ invariant one-forms $\omega^i(x)$, $i = 1, 2, 3$ which satisfy $d\omega^i = \epsilon_{jk}^i \omega^j \wedge \omega^k$ where ϵ_{jk}^i is the completely antisymmetric tensor of rank 3. In terms of Euler angle coordinates (ψ, θ, ϕ) on $SU(2)$ which take values in the range $0 \leq \psi \leq 4\pi$, $0 \leq \theta \leq \pi$, $0 \leq \phi \leq 2\pi$ (see also MTW (1973), p.808) we have

$$\begin{aligned} \omega^1 &= \cos \psi \, d\theta + \sin \psi \, \sin \theta \, d\phi \\ \omega^2 &= \sin \psi \, d\theta - \cos \psi \, \sin \theta \, d\phi \\ \omega^3 &= d\psi + \cos \theta \, d\phi \end{aligned} \quad (4)$$

Written out in terms of the coordinate differentials $d\psi$, $d\theta$, $d\phi$ we get for the line element of the spacetime metric (3)

$$\begin{aligned} ds^2 &= -dt^2 + c(t)^2 \, d\psi^2 + (a(t)^2 \cos^2 \psi + b(t)^2 \sin^2 \psi) d\theta^2 \\ &+ \{\sin^2 \theta (a(t)^2 \sin^2 \psi + b(t)^2 \cos^2 \psi) + c(t)^2 \cos^2 \theta\} d\phi^2 \\ &+ (a(t)^2 - b(t)^2) \sin 2\psi \, \sin \theta \, d\theta \, d\phi + 2c(t)^2 \cos \theta \, d\psi \, d\phi \end{aligned} \quad (5)$$

This is the toy-model spacetime metric (with $a(t), b(t), c(t)$, the three scale-functions, as degrees of freedom) which we want to evolve on approach to the “big crunch” space-time singularity where the three-volume of the metric collapses to zero.

The space is closed and the three-volume (Landau and Lifshitz p.390) of the compact space is given by

$$V = \int \sqrt{-g} \, d\psi \, d\theta \, d\phi = \int \sqrt{\gamma} \, \omega_1 \wedge \omega_2 \wedge \omega_3 = 16\pi^2 abc \quad (6)$$

When $a = b = c = R/2$ the space reduces to a space of constant positive curvature with radius of curvature $R = 2a$, which is the metric of highest symmetry on the group space $SU(2)$. The volume (6) then reduces to the three-volume $V = 2\pi^2 R^3(t)$ of the compact (isotropic) *Robertson-Walker* space. If we couple the gravitational field to perfect fluid matter, the cosmological model with the ansatz (3) for the metric is an anisotropic generalization of the well known compact FRW model: It has different “Hubble-constants” along different directions in the three-space. One may also interpret the metric (3) as a closed FRW universe on which is superposed circularly polarized gravitational waves with the longest wavelength that will fit into a closed universe (cf. D.H. King (1991) and references therein).

Since the metric (3) is spatially homogeneous, the full non-linear Einstein equations for the metric are a set of ordinary (non-linear) differential equations. To see this more explicitly, introduce in place of the quantities a, b, c , their logarithms

¹⁰If the metric is coupled to matter, e.g. perfect fluid matter, the assumption of diagonality of $\gamma_{ij}(t)$ is a *simplifying* ansatz. If no non-gravitational matter is present, the vacuum Einstein equations will automatically make the off-diagonal components of γ_{ij} vanish for a space with invariance group $G = \text{SU}(2)$, see e.g. Bogoyavlensky (1985), p. 34.

$\alpha = \ln a$, $\beta = \ln b$, $\gamma = \ln c$ and a new time variable $\tau = \int dt/abc$ in place of the proper (synchronous) time t appearing in the metric (3), cf. Landau and Lifshitz (1975), §116-119. With the inclusion of a perfect fluid matter source, the space-space components of Einstein's equations then reads

$$\begin{aligned} 2\alpha_{\tau\tau} &= \frac{d^2}{d\tau^2}(\ln a^2) = (b^2 - c^2)^2 - a^4 + 8\pi G(\rho - p) a^2 b^2 c^2 \\ 2\beta_{\tau\tau} &= \frac{d^2}{d\tau^2}(\ln b^2) = (c^2 - a^2)^2 - b^4 + 8\pi G(\rho - p) a^2 b^2 c^2 \\ 2\gamma_{\tau\tau} &= \frac{d^2}{d\tau^2}(\ln c^2) = (a^2 - b^2)^2 - c^4 + 8\pi G(\rho - p) a^2 b^2 c^2 \end{aligned} \quad (7)$$

and the time-time component reads

$$(\alpha + \beta + \gamma)_{\tau\tau} - 2(\alpha_\tau\beta_\tau + \alpha_\tau\gamma_\tau + \beta_\tau\gamma_\tau) = -4\pi G(\rho + 3p) a^2 b^2 c^2 \quad (8)$$

The quantities p and ρ denote the pressure and the energy density of the fluid. Adopting the standard ansatz that a barotropic equation

$$p = (\gamma - 1)\rho \quad (9)$$

(γ constant) relates the two quantities one may easily show that the equations of motion (7),(8) have a first integral

$$\tilde{I} = I - 8\pi G\rho a^2 b^2 c^2 = 0 \quad (10)$$

where

$$I = \alpha_\tau\beta_\tau + \alpha_\tau\gamma_\tau + \beta_\tau\gamma_\tau - \frac{1}{4}(a^4 + b^4 + c^4) + \frac{1}{2}(a^2 b^2 + a^2 c^2 + b^2 c^2) \quad (11)$$

To be in accordance with the full set of Einstein equations the solution should have $\tilde{I} = 0$. The dynamical equations for the compact FRW cosmology is recovered in the case of $a = b = c = R/2$.

Important astrophysical examples of the barotropic equation (9) are $\gamma = 1, 4/3, 2$ corresponding to the cases of "dustlike" matter, "radiation" matter and "stiff matter" respectively. The energy density scales with the volume of the space as $\rho \sim V^{-\gamma}$, see also e.g. Landau and Lifshitz (1975) or Kobl and Turner (1990).

One may show (cf. also Landau and Lifshitz (1975), p.390) that sufficiently near the singularity the perfect fluid matter terms (appearing on the right hand side of the equations (7) and (8)) may be neglected if the equation of state $p \leq 2/3\rho$. Thus, it is sufficient to investigate the "empty space equation" (the vacuum Einstein equations) $R_{\mu\nu} = 0$ even if "dust" ($\gamma = 1$ in equation (9)) and "radiation" fluids ($\gamma = 4/3$) are included in the mixmaster big crunch collapse: One says that the mixmaster cosmology is "curvature dominated" in the region sufficiently near the space-time singularity! In physical terms this means that sufficiently near the space-time singularity the self-interaction of the gravitational field completely dominates the dynamical evolution and contributions from (non-gravitational) matter may be neglected in the study. This conclusion clearly does not apply to the case of a "stiff matter" fluid (where $p = \rho \sim (abc)^{-2}$) and - of course - does not a priori apply to sources of other physical origin. Such other material sources (non-Abelian Yang Mills fields, etc.) might very well couple to and significantly alter the dynamical structure of the gravitational collapse.

Note, that in the reversed time-direction, when the mixmaster spacetime metric (3) is evolved away from its singularity (i.e. as the volume $V = 16\pi^2 abc$ of the space increases) the matter terms gradually become more important and eventually dominate the dynamical evolution of the mixmaster metric (3) and the matter terms may lead to isotropization (though, not fast enough to explain the observed isotropy today, cf. e.g. Doroshkevich, Lukash and Novikov (1973) and Lukash (1983)).

Despite the fact that this metric is widely known, it is remarkable that only quite recently (cf. X. Lin and R.M. Wald (1990)) it has been rigorously proven to recollapse (this is also true in the “vacuum” case, i.e. when no perfect fluid, or any other matter source, is included in the model). So, we know that the mixmaster spacetime metric has *two* spacetime singularities (like the compact FRW cosmology): A “big bang” and a “big crunch”!

2.2. The three-volume of the mixmaster space-time metric cannot oscillate if evolved according to the vacuum Einstein equations

The three-volume $V = 16\pi^2 abc$ of the model-universe cannot oscillate. This can be shown in several (not truly independent) ways:

One may derive this fact directly from the governing set of differential equations (see S.E. Rugh (1990a) and S.E. Rugh and B.J.T. Jones (1990)). We note, that statements on monotonicity of the three-volume are equivalent whether given in t or in τ time, since $dt = abc d\tau$ and $abc > 0$. The property of $\ln V$ being a *concave* function (negative second derivative) does not translate from t to τ time. Below we show that $\ln V$ is a concave function in the t time variable (but not in τ time).

Neglecting, for notational convenience, the factor $16\pi^2$ in the expression for the three-volume, we have $\ln V \equiv \ln a + \ln b + \ln c \equiv \alpha + \beta + \gamma$, and the R_{00} equation for the mixmaster metric reads

$$\frac{1}{2}(\alpha + \beta + \gamma)_{\tau\tau} \equiv \frac{1}{2}(\ln V)_{\tau\tau} \equiv \alpha_\tau \beta_\tau + \alpha_\tau \gamma_\tau + \beta_\tau \gamma_\tau \quad (12)$$

From the definition $\partial_t = (abc)^{-1} \partial_\tau = V^{-1} \partial_\tau$ one arrives at $\partial_t^2 = V^{-2} \{\partial_\tau^2 - (\ln V)' \partial_\tau\}$ and hence

$$\begin{aligned} V^2 \partial_t^2 (\ln V) &= V \partial_t^2 V - (\partial_t V)^2 = (\ln V)'' - ((\ln V)')^2 \\ &= 2(\alpha' \beta' + \alpha' \gamma' + \beta' \gamma') - (\alpha' + \beta' + \gamma')^2 \\ &= -\alpha'^2 - \beta'^2 - \gamma'^2 \leq 0 \end{aligned} \quad (13)$$

It follows that $(\ln V)$ - and therefore the volume itself, V , can have *no local minimum* (where we should have $\dot{V} = 0, \ddot{V} > 0$). As a corollary it follows that volume oscillations are not possible: After a transient (eventually passing a maximum in the three-volume), the volume of the three-space should decrease *monotonically* as the mixmaster gravitational collapse evolves towards the “big crunch” singularity.

We may, also, consider the Raychaudhuri equation, A. Raychaudhuri (1955), which governs an equation for the relative volume expansion $\Theta = \partial_t(\log \{Volume\})$ with respect to the coordinate t time. The equation was independently discovered by Lev Landau and A. Raychaudhuri (see S.W. Hawking & G.F.R. Ellis (1973), p.84) and is derived from the Einstein equations for a spacetime metric in a synchronous reference

frame coupled to co-moving perfect fluid matter. The general form of Raychaudhuri's equation is:

$$\begin{pmatrix} \text{expansion} \\ \text{derivative} \end{pmatrix} = - \begin{pmatrix} \text{energy density} \\ \text{term} \end{pmatrix} - \begin{pmatrix} \text{shear} \\ \text{term} \end{pmatrix} - \begin{pmatrix} \text{expansion} \\ \text{term} \end{pmatrix} + \begin{pmatrix} \text{vorticity} \\ \text{term} \end{pmatrix}$$

or

$$\dot{\Theta} = -R_{\mu\nu}u^\mu u^\nu - \sigma^2 - \frac{1}{3}\Theta^2 + 2\Omega^2 . \quad (14)$$

A dot denotes derivation with respect to the time t . The term $R_{\mu\nu}u^\mu u^\nu = 4\pi G(\rho + 3p)$ refers to the *co-moving* perfect fluid source (the nongravitational matter) included in the cosmological model, whereas σ^2 and Ω^2 denote the "shear" and "vorticity" scalars contracted from the shear and vorticity tensors of the metric field (see, also, S.W. Hawking and G.F.R. Ellis (1973)). One may show that the metric (3) has no vorticity $\Omega^2 = 0$. For the quantity

$$\mathcal{G} \equiv \sqrt[3]{abc} \propto \sqrt[3]{\text{Volume}} ,$$

which is related to the relative volume expansion Θ as

$$\Theta = \frac{d/dt \{\text{Volume}\}}{\text{Volume}} = 3\frac{\dot{\mathcal{G}}}{\mathcal{G}} , \quad \dot{\Theta} + \frac{1}{3}\Theta^2 = 3\frac{\ddot{\mathcal{G}}}{\mathcal{G}} ,$$

one obtains, after a little algebra, the equation

$$\frac{\ddot{\mathcal{G}}}{\mathcal{G}} = \frac{d^2/dt^2(\sqrt[3]{abc})}{\sqrt[3]{abc}} = -\frac{4\pi G}{3}(\rho + 3p) - \frac{2}{3}\sigma^2 \leq 0 . \quad (15)$$

Thus, provided ρ and p are not negative, $\ddot{\mathcal{G}}$ is always non-positive implying that \mathcal{G} cannot have a local minimum (where one should have $\dot{\mathcal{G}} = 0$ and $\ddot{\mathcal{G}} > 0$). I.e., according to Raychaudhuri's equation, the three-volume for any cosmological model without vorticity $\Omega = 0$ (our diagonal mixmaster toy model collapse (3) belongs to this class) cannot be oscillatory, but can only have one maximum like the FRW cosmology. In the empty case, $p = \rho = 0$, the conclusion applies equally well.

2.3. Volume oscillations as a "probe" on numerical solutions

The fact that solutions to the vacuum Einstein equations should not have oscillating three volumes is no surprise. The idea, however, is to use the property of "no oscillations" as a "probe" to examine the validity of some previous investigations which were done on this model. In the references Zardecki (1983), Francisco and Matsas (1988), and in fact in Barrow (1982, 1984) and Barrow and Silk (1984), the volume behavior of the depicted evolutions is not in accordance with the conclusion arrived at here. They have such oscillations and can therefore not be proper solutions in agreement with the Einstein equations for positive or zero non-gravitational energy densities. In fact, some of these models effectively included *stiff* matter with *negative* mass densities. (cf. S.E. Rugh (1990a) and S.E. Rugh and B.J.T. Jones (1990)). That is, one may easily show¹¹

¹¹In the case of "stiff matter", $p = \rho$, i.e. $\gamma = 2$ in (9), and we have $\rho \sim V^{-2}$. The 11-, 22-, 33-components of the Einstein field equations (7) attain the same form as the vacuum equations. The exception is the 00-component of the Einstein equations (8) and thereby the first integral constraint (10). Since the energy density scales with the three volume as $\rho \sim V^{-2}$ we put $\rho = \rho_0(abc)^{-2}$ which gives $I = 8\pi G\rho_0$. For the vacuum Einstein equations we ought to have $I \equiv 0$. However, if initial data fail to satisfy the zero density constraint, the corresponding solutions act as if we had included "stiff matter" with $p_0 = \rho_0 = (8\pi G)^{-1}I \neq 0$.

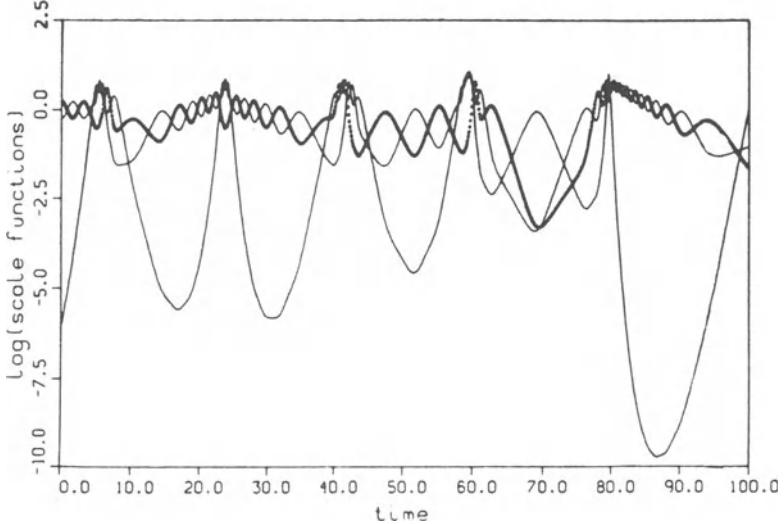


Figure 1. A typical evolution of the logarithmic scale functions $\alpha = \log a$, $\beta = \log b$, $\gamma = \log c$ (the thick curve is $\alpha = \log a$) as a function of τ -time, $\tau = \int dt/abc$, is rather interesting if we select initial conditions which have $I < 0$ and thus effectively introduces stiff matter with negative energy density. Apparently the evolution of the scale functions of the metric (3) is highly irregular.

that vacuum solutions which fail to satisfy the first integral constraint $I = 0$, with I given in (11), correspond to the inclusion of “stiff matter” (i.e. a perfect fluid with equation of state “ $p = \rho$ ”) with energy density

$$p = \rho = (8\pi G)^{-1} \frac{I}{a^2 b^2 c^2} .$$

The character of the solutions depends on the sign of I .

For $I < 0$ the negative mass densities create a *repulsion effect*, which causes oscillations in the three volume $V = 16\pi^2 abc$. In such solutions a typical evolution of the parameters $a(t)$, $b(t)$, $c(t)$ of the metric (3) will be like in fig. 1 where the scale functions - or rather their logarithm’s $\alpha = \log a$, $\beta = \log b$, $\gamma = \log c$ - are plotted against the standard (Landau and Lifshitz, Vol. II, §118) time variable $\tau = \int dt/abc$.

One may also display the “degree of anisotropy” of the spacetime metric in the anisotropy variables (ADM variables, see later)

$$\vec{\beta} = (\beta_+, \beta_-) = \left(\frac{1}{6} \log\left(\frac{ab}{c^2}\right), \frac{1}{2\sqrt{3}} \log\left(\frac{a}{b}\right) \right). \quad (16)$$

$\vec{\beta} = \vec{0}$ means no anisotropy, while a huge $|\vec{\beta}|$ means that our model (this is the case for proper solutions near the singularity, cf. fig. 5) is very anisotropic. In terms of this parameterization the metric (3) reads, cf., e.g., MTW (1973),

$$ds^2 = -dt^2 + e^{-2\Omega} (e^{2\beta})_{ij} \omega^i(x) \omega^j(x) \quad (17)$$

where $\Omega = -\frac{1}{3} \log(abc)$ and β_{ij} is the traceless, diagonal matrix with the diagonal elements $\beta_+ + \sqrt{3}\beta_-$, $\beta_+ - \sqrt{3}\beta_-$ and $-\beta_+$. The evolution of the metric is decomposed in *expansion* (volume change parametrized by Ω) and *anisotropy* (shape change

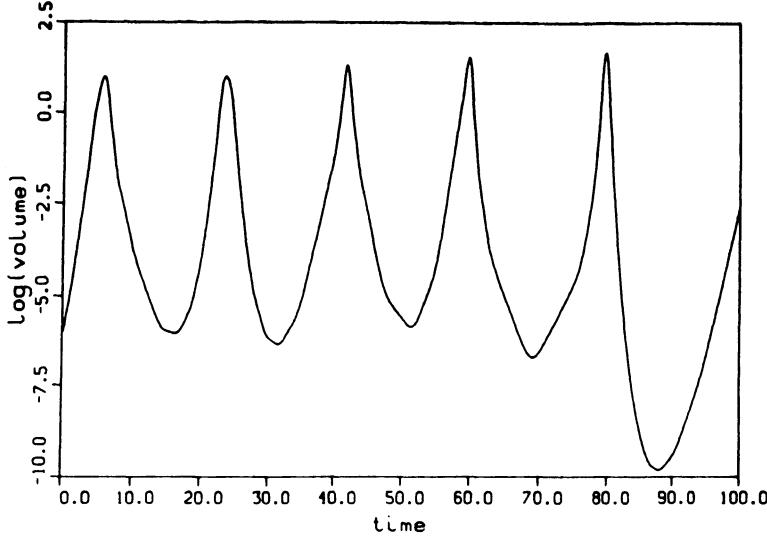


Figure 2. Behavior of the logarithm of the three-volume $V = 16\pi^2abc$ as a function of the τ -time $\tau = \int dt/abc$ corresponding to the behavior of the scale functions in fig.1. Due to the effective inclusion of negative energy densities (see text) the solutions do not satisfy the reasonable conditions on the energy momentum tensor required for the Hawking-Penrose singularity theorems to apply. The metric does not evolve towards a singularity: The negative mass density creates a repulsion effect and, as a result, the three-volume oscillates in time.

parametrized by $\vec{\beta} = (\beta_+, \beta_-)$). The trajectories of the “anisotropy” of the (erratic volume oscillating) mixmaster model, which correspond to the sketched solutions above, fig.1, is displayed in fig.3 (the C_3 , symmetry is apparent and is of course expected from the symmetry under the interchange $a \leftrightarrow b \leftrightarrow c$ of the scale factors in the metric (3)).

The mixmaster equations pass the Painlevé test

It is interesting that Contopoulos *et al* (1993) recently have performed a Painlevé analysis on the set of mixmaster equations (7), for $\rho = p = 0$, and find that the set of mixmaster equations pass the Painlevé test. Apparently, this analysis does not utilize the additional information from the first integral constraint $I = 0$. Thus the equations of motion for the mixmaster space-time metric also pass the Painlevé test for the case $I < 0$ which, according to Contopoulos *et al*, is a strong indication that the trajectories corresponding to fig.1 and fig.3 (as well as fig.4 and fig.5 for $I = 0$) are integrable, i.e. that two additional *constants of motion (symmetries)* besides the Hamiltonian can be found. If an integrable system could produce an orbit like the one in fig.3 it would be surprising. We have previously searched for such additional integrals.¹² Our results, so far, indicate the lack of existence of such additional integrals in the equations of motion for the case $I < 0$.

2.4. The mixmaster gravitational collapse evolved towards the spacetime singularity under the governing vacuum Einstein equations

The dynamics of the spacetime metric (3) will be very complicated (though not as complicated as in figures 1,2 and 3 above) when evolved according to the vacuum Einstein

¹²F.Christiansen, H.H.Rugh and S.E.Rugh, unpublished investigations

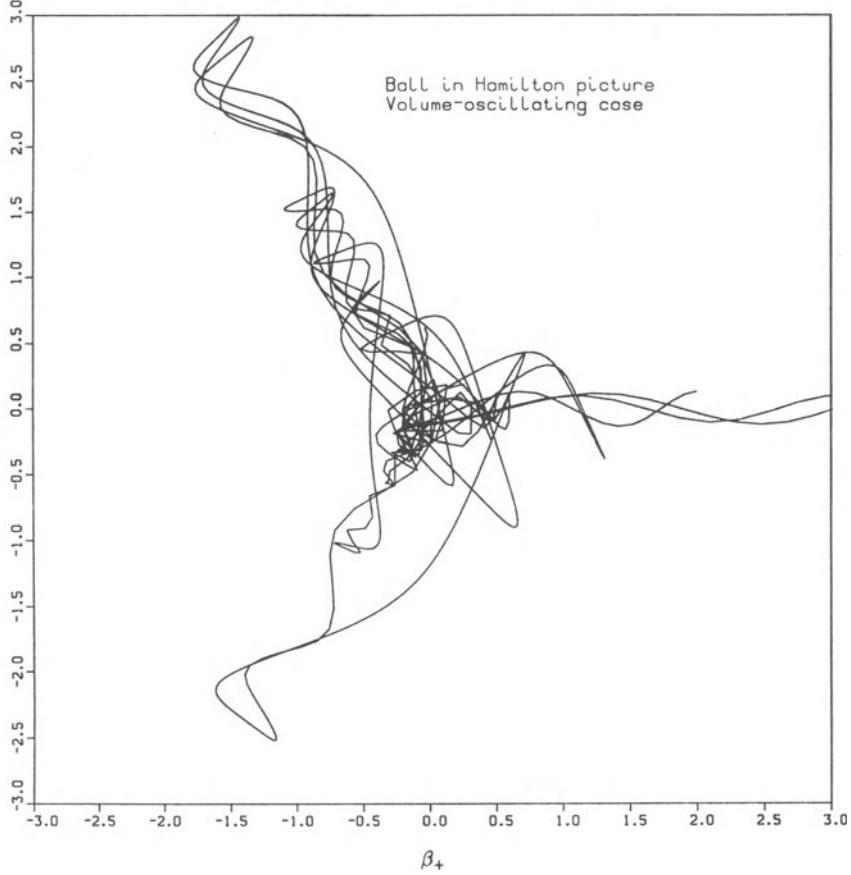


Figure 3. Solutions corresponding to fig.1 but mapped out in the anisotropy variables $\vec{\beta} = (\beta_+, \beta_-)$. Not surprisingly, this behavior is reflected in positive values of the maximal characteristic Lyapunov exponent.

equations $R_{\mu\nu} = 0$. The deterministic evolution of the scale factors $a(t)$, $b(t)$, $c(t)$ and their first derivatives is described by the set of six coupled, first order, differential equations (7) (we put $p = \rho = 0$) constrained by the first integral (10). Due to the scale invariance of the Einstein equations there are four degrees of freedom in the problem. We may distinguish between solutions which are axisymmetric (an integrable case) and solutions without any axisymmetry.

With rotational invariance about one axis¹³ ($a=b$, say) in the SU(2) homogeneous 3-space, we yield for the metric (3) the special case of the Taub spacetime metric of the form

$$ds^2 = -dt^2 + a^2(t)((\omega^1)^2 + (\omega^2)^2) + c^2(t)(\omega^3)^2 ,$$

and in this axisymmetric case the vacuum equations admit the *exact* solution (given in A.H.Taub (1951))

$$a^2 = b^2 = \frac{p \cosh(2p\tau + \delta_1)}{2 \cosh^2(p\tau + \delta_2)} , \quad c^2 = \frac{2p}{\cosh(2p\tau + \delta_1)} , \quad (18)$$

¹³Note that the FRW solution, being invariant under rotations about any axis in the SU(2) homogeneous 3-space (i.e. isotropic) is not obtained from the Taub solution by putting $a=b=c$. The Taub solution is a *vacuum* solution whereas the FRW solution is a solution with matter (a perfect fluid) in the model.

where p, δ_1, δ_2 are constants. However, this axisymmetric solution is *unstable* with respect to small perturbations in the parameter space of scale functions (a, b, c, \dots) , see also C.W.Misner (1969). For a numerical investigation of the Taub solutions, see C. Behr (1962).

The dynamical behavior for a typical mixmaster gravitational collapse (without additional symmetries) is displayed in fig.4 and fig.5 where the three axes (a, b, c) are followed as a function of the standard Landau and Lifshitz time coordinate $\tau = \int dt/abc \rightarrow \infty$. Near the singularity where the scale functions collapses to zero $a, b, c \rightarrow 0$ we prefer to display the logarithm $\alpha = \log a, \beta = \log b, \gamma = \log c$ of the scale functions. The “anisotropy” of the toy-model grows without limit on approach to the big crunch singularity¹⁴

We have selected a set of reference initial conditions as in A. Zardecki (1983), but have adjusted the value of c' to make the first integral vanish to *machine precision*. (Such an adjustment is indeed necessary. Cf. discussions in S.E. Rugh (1990a) and D. Hobill *et al* (1991)). This yields the starting conditions

$$\begin{aligned} a &= 1.85400.., b = 0.438500.., c = 0.085400.. \\ a' &= -0.429200.., b' = 0.135500.., c' = 2.964843279..... \end{aligned} \quad (19)$$

All integrations have been performed by a standard *fourth order Runge Kutta* algorithm, and each calculation takes less than one CPU-minute.

We distinguish between two qualitatively different phases of the evolution of the spacetime metric (3) towards the “big crunch” singularity:

Transient behavior: The model cosmology trajectory “catches up” its initial conditions, and eventually passes a maximum in three-volume. The volume then begins to decrease monotonically. (Fig.4)¹⁵

Asymptotic behavior near the singularity: The evolution towards the “big crunch” singularity is fast attracted¹⁶ into an infinite sequence of oscillations of the three scale functions (fig.5). We may identify the Kasner segments (Kasner epochs) between each “bounce”, and the combinatorial model by Belinskii *et al* describes well the transitions from one Kasner segment to the next. (Table 1).

Extracting lower dimensional signals from the gravitational collapse

In an ever expanding phase space (cf., also, discussions later) it is natural to try to “project out” some lower dimensional (compact) signal¹⁷, say, which captures some of

¹⁴If the anisotropy-variables $\vec{\beta}$, calculated according to (16), are mapped out corresponding to the long τ time behavior of the α, β, γ variables depicted in fig.5, they will be pieces of straight line segments corresponding to bounces from outward expanding (almost infinitely steep) potential walls. It is like a game of billiards played on a triangular shaped table with outward expanding table boundaries, with a ball which moves faster than the outward expanding boundaries (cf., also, displayed figures of this behavior in S.E. Rugh (1990a) and B.K. Berger (1990)).

¹⁵If we want a more realistic model cosmology at this stage, matter terms should be included here, since the omission of matter contributions for the mixmaster toy-model cosmology is only justified when we are sufficiently near the “big crunch” singularity.

¹⁶Note that the sequence of oscillations of the scale functions, as shown in fig.5, is an attractor for almost any initial condition, but not initial conditions with axisymmetry.

¹⁷It is not uncommon to “probe” chaos in dynamical flows which take place in higher dimensional phase spaces by extracting lower dimensional time-signals from the flow. (E.g. probing turbulence and complexity in Navier-Stokes hydrodynamical flows by measuring time signals of a temperature probe, placed at some given space point in the fluid). Extraction of a single physical variable $\xi(t)$ with “chaotic” behavior (a “time” signal or a discrete map ξ_n) mirrors “chaos” in the full phase flow.

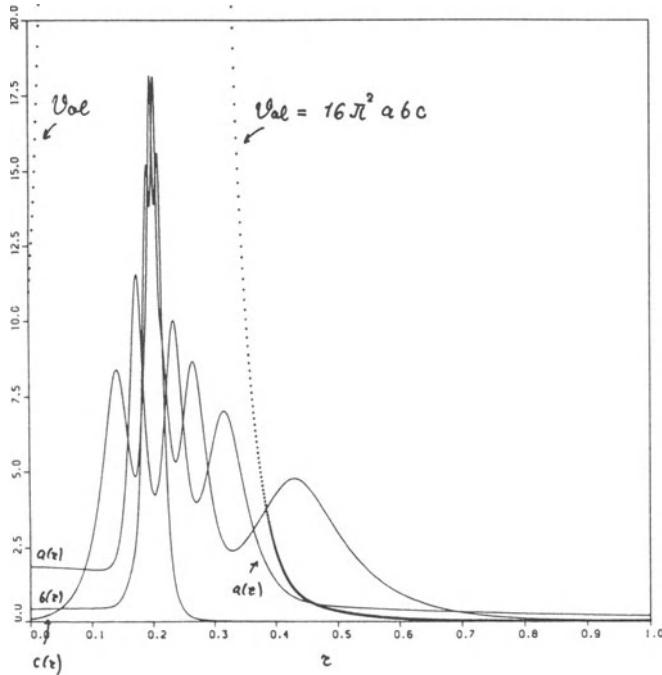


Figure 4. Transient evolution of the three scale functions of the mixmaster space-time metric governed by the *vacuum* Einstein equations with starting conditions (19). We have displayed the “*transient*” τ time interval $0 \leq \tau = \int dt/abc \leq 1$. (From S.E. Rugh (1990 a)). The three-volume has a maximum at $\tau \approx 1/5$ and then decreases monotonically (turning into an oscillatory behavior of the scale functions as shown in fig.5) towards the big crunch singularity. (It is interesting, and surprising, if this displayed behavior should turn out to be integrable, as suggested by Contopoulos *et al* (1993)).

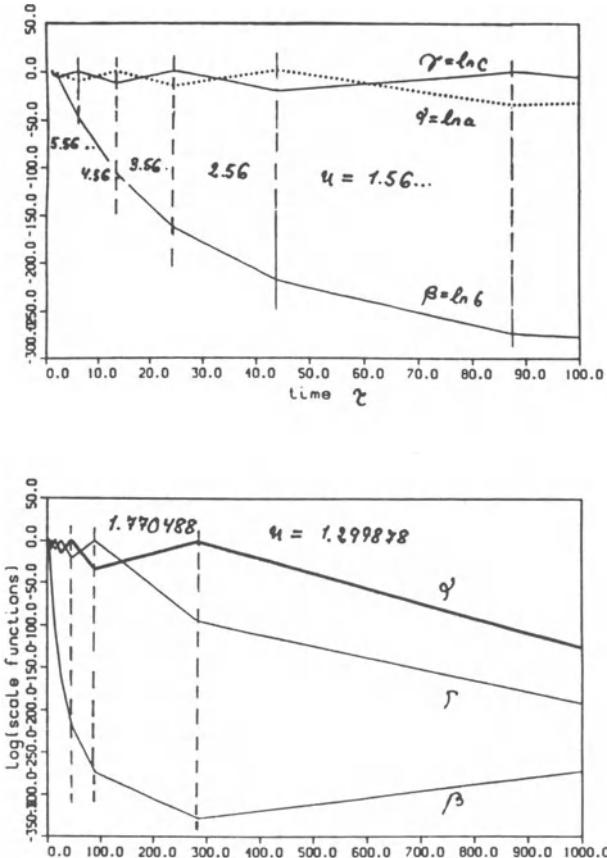


Figure 5. Asymptotic evolution of the three scale functions (typical figures from a numerical experiment described in S.E. Rugh (1990a)): The axes $\alpha = \ln a$, $\beta = \ln b$, $\gamma = \ln c$ of the toy-model gravitational collapse, given by the metric (3), is followed in τ -time ($\tau = \int dt/abc$) towards the “big crunch” singularity under the governing *vacuum* Einstein equations. The spacetime singularity is reached in *finite* t time, but in τ time the dynamical evolution is stretched out to infinity $\tau \rightarrow \infty$. The three-volume of the space shrinks monotonically to zero, the anisotropy of the model grows without limit and the Weyl curvature $C^2 = C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}$ also diverges on approach to the spacetime singularity.

the recurrent “chaotic” properties of the model. The dynamical behavior of the toy-model gravitational collapse is indeed in certain aspects chaotic, as captured by the parameter

$$u = \frac{\min\{\alpha', \beta', \gamma'\}}{\alpha' + \beta' + \gamma' - (\min\{\alpha', \beta', \gamma'\} + \max\{\alpha', \beta', \gamma'\})} \quad (20)$$

Extracting this parameter “u” (the so-called “Lifshitz-Khalatnikov” parameter “ $u \in \mathbf{R}$ ”, see also Landau and Lifshitz, §116-119) from the displayed trajectories we get table 1. We find that the gravitational collapse is *extraordinarily well* described by the *BKL*-combinatorial model which is summarized below, following O.I.Bogoyavlensky and S.P.Novikov (1973). “*BKL*” refers to the originators of this combinatorial description of the behavior of the scale functions: V.A. Belinskii, I.M. Khalatnikov and E.M. Lishitz. See also I.M. Khalatnikov et al. (1985) and references therein.¹⁸

Table 1. This table summarizes the evolution of numerically extracted values of the Lifshitz-Khalatnikov parameter “u” corresponding to solutions of the vacuum Einstein equations as those depicted in fig.5. The description of the evolution of the “u” parameter in terms of the “Farey map” or “Gauss map” is in complete agreement with this table. From S.E. Rugh (1990a), p.98. (The τ -time values offered in the table correspond to τ -times when the collapse orbit is well beyond a bounce and has reached a new straight line behavior (“Kasner epoch”) in fig.5.

$\tau = \int dt/abc$	u	$1/u$
1	5.564816	
6	4.564816	
16	3.564815	
27	2.564816	
48	1.564816	
96	1.770488	0.5648160
305	1.297878	0.7704884
4700	3.357077	0.2978782
50000	2.357077	
180000	1.357077	
500000	2.800516	0.3570771
3000000	1.800510	
19000000	1.249229	0.8004938
75000000	4.013535	0.2491569

With the parametrization

$$(p_1(u), p_2(u), p_3(u)) = (-u, 1+u, u(1+u))/(1+u+u^2)$$

¹⁸The accordance with the *BKL*-combinatorial model for the dynamical evolution of the mixmaster collapse has been further investigated in the work of B.K. Berger, cf., e.g., Berger (1993). For a description of the more complete 4-parameter map, derived by Belinskii *et al*, see also D. Chernoff and J.D. Barrow (1983).

the *BKL*-piecewise approximation of the scale functions a, b, c by the power law functions (Kasner epochs) $t^{p_1}, t^{p_2}, t^{p_3}, p_i = p_i(u)$, is (on approach to the spacetime singularity) described by the sequence of states

$$(u_0, \sigma_0) \rightarrow (u_1, \sigma_1) \rightarrow (u_2, \sigma_2) \rightarrow \dots \quad (21)$$

where the *BKL*-“Kasner state” transformation (“alternation of Kasner epochs”) is given by

$$\begin{cases} (u, \sigma) \rightarrow (u - 1, \sigma\sigma_{12}) & (2 \leq u < \infty) \\ (u, \sigma) \rightarrow (1/(u - 1), \sigma\sigma_{12}\sigma_{23}) & (1 \leq u \leq 2) \end{cases} \quad (22)$$

the “Kasner state” being described by the pair

$$(u, \sigma) ; \sigma = \begin{pmatrix} 1 & 2 & 3 \\ i & j & k \end{pmatrix}, \sigma_{12} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \sigma_{23} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad (23)$$

and σ denoting the permutation of the three Kasner axes.

2.5. The (compressed and stretched) astronomer who falls into the space-time singularity

A ‘*freely falling astronomer*’ who falls into the mixmaster spacetime singularity will experience a tidal field, in which he is *compressed* along two directions and *stretched* (expanded) in one. The directions of these gravitationally induced stresses are permuted infinitely many times (in a not-predictable way!) on approach to the space time singularity. A picture in Kip S. Thorne (1985) sketches the tidal gravitational forces felt by an observer (an astronomer) who falls into the singularity. Such tidal forces are produced by spacetime curvature. The astronomer, who in this example makes up the tidal field instrument (see, e.g., MTW, p.400-404), feels, in his local inertial frame, tidal accelerations given by the equation of geodesic deviation, $d^2\xi_j/dt^2 = R_{joko} \xi^k$, where ξ is the separation vector between two freely falling test particles (two reference points in the body of an astronomer, falling freely along geodesics - if we neglect internal elastic forces in the body of the observer (justified, if the spacetime curvature is big?). There is in the local inertial frame of the infalling observer a preferred choice of coordinate axes ($i = 1, 2, 3$) which diagonalizes the tidal field. In terms of these “principal axes” the component R_{ioio} produces, according to the equation of geodesic deviation, $\ddot{\xi}_i/\xi_i = -R_{ioio}$, a tidal *compression* or *stretching* along direction “ i ”, depending on whether R_{ioio} is positive or negative.

One may swindle a bit and write down the Riemann curvature components of the Kasner-segments as if they were given by the Riemann curvature of the Kasner metric (instead of the Riemann tensor components of the spacetime metric (3)). The Riemann tensor components of the Kasner metric reads $R_{ioio} = -p_i(p_i - 1)t^{-2}$, $i = 1, 2, 3$, and we get, in terms of the Lifshitz-Khalatnikov parameter $u \in \mathbf{R}$, the following expressions for the tidal stresses:

Along the two axes of compression

$$\frac{\ddot{\xi}}{\xi} = -\frac{u(u + 1)}{(1 + u + u^2)^2} t^{-2} \text{ resp. } \frac{\ddot{\xi}}{\xi} = -\frac{u^2(u + 1)}{(1 + u + u^2)^2} t^{-2} .$$

Along the axis of expansion

$$\frac{\ddot{\xi}}{\xi} = +\frac{u(u + 1)^2}{(1 + u + u^2)^2} t^{-2} > 0 .$$

The tidal stresses grow up like $\sim t^{-2}$ where t denotes the *finite* time distance (as measured by the infalling observer) until the spacetime singularity is reached. The successive shifts $u \rightarrow u - 1$ & C in the parameter “ u ”, are governed by the combinatorial model (22). During each “era” (comprised of cycles $u \rightarrow u - 1$ until u reaches the value $1 \leq u \leq 2$) one of the principal axes experiences a continual tidal compression while the other two principal axes oscillate between compression and stretch. At a change of “era” (i.e. when $u \rightarrow 1/(u - 1)$ for $1 \leq u \leq 2$) there is a change in the role of the axes (another of the three principal axes experiences a continual tidal compression while the other two axes oscillate). Note, that “real” *physical* quantities, measurable by *tidal field instruments* exhibit chaotic oscillations which condense infinitely on approach to the spacetime singularity at $t \rightarrow 0$.

The astronomer may feel a little “worried”, being compressed like this in two directions and stretched (expanded) in one. The directions of these gravitationally induced stresses are even permuted infinitely many times on approach to the spacetime singularity.

2.6. The “Farey map” with “strong intermittency” encodes the “per bounce” dynamics of the mixmaster gravitational collapse

The *unpredictability* of the mixmaster gravitational collapse does not originate from the oscillations of the scale functions (as described by $u \rightarrow u - 1$) *within* a given major cycle (an “era” of oscillations), but rather from the *shifts* between major cycles (described by $u \rightarrow 1/(u - 1)$, $1 \leq u \leq 2$). These shifts give rise to the (highly chaotic) Gauss map, cf. e.g. J.D. Barrow (1982). The Gauss map is well known from “chaos theory” (cf., e.g., R.M. Corless *et al* (1990) and references therein) and acts as a *left shift* on the continued fraction representation of numbers on the unit interval. The Gauss map has positive Kolmogorov entropy, $h = \pi^2/6(\ln 2) \gg 0$ and it has the Bernoullian property¹⁹ and, in that sense, it is as random as that of flipping coins (or a roulette wheel).

A map, which describes the “per bounce” evolution of u (and not, merely, the “per major cycle” evolution of the u parameter) is easily found²⁰ and is known as the “Farey map”: Make the substitution $u = 1/x$ in (22) and write this map in terms of the parameter $x \in]0, 1]$. This yields the following map on the unit interval (the “Farey map”²¹)

$$x \rightarrow \mathcal{F}(x) = \begin{cases} x/(1-x) & \text{if } 0 \leq x \leq 1/2 \\ (1-x)/x & \text{if } 1/2 \leq x \leq 1 \end{cases}$$

Whereas the Gauss map has an *infinity* of branches (cf., e.g., pictures in J. Barrow (1982) or R.M. Corless *et al* (1990)) the Farey map has only two branches (the left, $x \leq 1/2$, and the right) and thus a natural *binary* symbolic dynamics (cf. sec.3.4) with a binary alphabet (which corresponds more directly to a symmetry-reduced binary symbolic dynamics encoding of the geodesic motion on the triangular billiard on the Poincaré disc).²² The Farey map contains the Gauss map as a sub-map, since one iteration of

¹⁹The Gauss map is Bernoullian in the sense that it may be extended to a two-dimensional invertible map which is isomorphic to a Bernoulli shift with an infinite alphabet.

²⁰H.H. Rugh and S.E. Rugh (1990), unpublished. See also B.K. Berger (1991).

²¹cf., e.g., M. Feigenbaum (1988) or Artuso, Aurell and Cvitanovic (1990), p.378, and references therein.

²²The Farey/Gauss map is closely related to the symbolic description of geodesic flows on so-called modular surfaces (found by Artin) on the Poincaré disc, see e.g. T. Bedford *et al* (1991), and therefore

the Gauss map corresponds to a transition from the right branch via oscillations in the left branch back to the right branch.

As regards the evolution of the Lifshitz-Khalatnikov parameter “ u ”, the Farey map takes into account both chaotic and “non-chaotic” segments of the one-perturbation BKL-combinatorial model for the gravitational collapse. The watch of the “Farey map” ticks one step forward (one iteration of the map) for each bounce against a wall, i.e. for each oscillation of the scale functions.

The Farey map has a *marginally stable* fixed point at the left end ($|\mathcal{F}'(0)| = 1$). This has an important influence on the instability properties of the map. Intuitively, the marginally stable point of the Farey map at $x = 1/u = 0$ corresponds to major cycles, containing an “infinite” number of oscillations (governed by the $u \rightarrow u - 1$ rule), i.e., trajectories which penetrate deeply into one of the three corner channels.

In general, a map \mathcal{F} with the asymptotic expansions $\mathcal{F}(x) \approx x + ax^\zeta + \dots$ (towards the marginally stable point at $x = 0$) and $\mathcal{F}(x) \approx 1 - b|1 - 2x|^{1/\alpha} + \dots$ (around the top point $x = 1/2$) will have an invariant measure $\mu(x)$, which has the asymptotic behavior²³

$$\mu(x) \sim x^\eta ; \quad \eta = \alpha + 1 - \zeta$$

near the origin $x = 0$. The situation $\eta > 0$ is then referred to a “weak intermittency” (there is a smooth invariant measure and the Lyapunov exponent is positive, $h = \lambda = \int \ln |\mathcal{F}'(x)| \mu(x) dx > 0$) while $\eta < 0$ implies “strong intermittency” (there is no normalized, smooth measure and $h = \lambda = 0$). As regards the Farey map we have $\alpha = 1, \zeta = 2$; i.e. $\eta = 1 + 1 - 2 = 0$ which is the borderline case. However, the Farey map is an example of strong intermittency. By direct inspection one may verify that $\mu(x) = 1/x$ is an invariant measure. This measure is non-normalizable, but is normalized to the δ -function measure $\delta(0)$ at the marginally stable fixpoint $x = 0$. Hence, all the measure is concentrated at the *stable* fixpoint $x = 0$ and the averaged Lyapunov exponent (the K-entropy) is zero. Although the overall Lyapunov exponent (Kolmogorov entropy) of the map is zero, all periodic orbits (which do not include $x = 0$) are unstable and have $\lambda > 0$.

This strong intermittency means that the stability properties are completely dominated by the “infinite number of corner oscillations” fixpoint at $x = 0$. However, we note that all this is within the *one-perturbation* combinatorial model (22) for the gravitational collapse, and therefore it is not valid in the regime $x = 1/u \approx 0$. Thus, there is a “cut off” towards the left end of the Farey map and the good ergodic properties are regained. A *two-perturbation* analysis has to take over and show that the trajectory then will leave the corner.²⁴

The binary “Farey tree”. Self-similarity of the collapse dynamics.

The near-singularity dynamics of the gravitational collapse, as regards the parameter $x = 1/u$, can be symbolically represented by the so-called “Farey tree”, which is a

a very nice connection exists (cf. also J. Pullin (1990)) between the Farey/Gauss map encoding of the mixmaster gravitational collapse and the description of the mixmaster collapse orbit on the Poincaré disc.

²³See, e.g., Z. Kaufmann and P. Szépfalusy (1989) or P. Szépfalusy and G. Györgyi (1986).

²⁴The avoidance of the so-called “dangerous case of small oscillations” (i.e. dangerous for the one-perturbation treatment in the BKL-combinatorial model) has been discussed in the work by Belinskii et al., cf. Khalatnikov et al. (1985) and references therein.

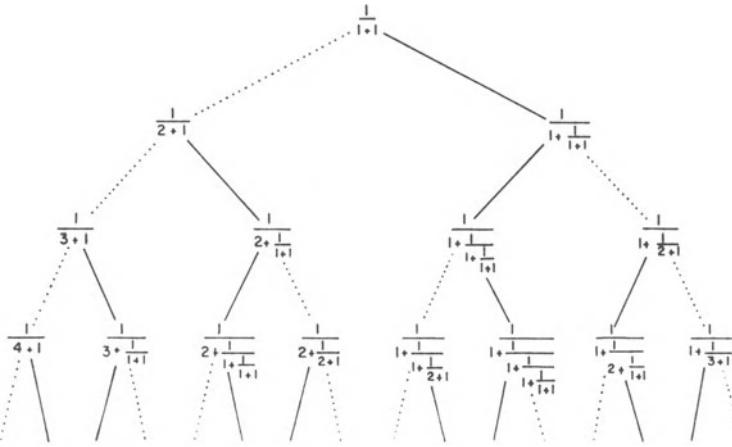


Figure 6. A continued fraction representation of the binary Farey tree (encoding the bounce dynamics of the gravitational collapse). The dotted lines continue a given branch (and correspond to repeated bouncing between the same two potential walls) while the full lines make a “zig-zag” movement down the tree (which corresponds to a transition to a third wall). The Farey tree rationals can be generated by backward iterates of the number $1/2$ by the Farey map, or, what is the same, by interpolating the rationals downwards by means of the “Farey mediants”. Figure from Cvitanovic and Myrheim (1989).

construction associated with the Farey map. The Farey tree is a “tree of rational numbers” (the Farey numbers). In fig.6 each Farey number has been represented by its continued fraction. Originally, the tree is constructed by starting with the endpoints of the unit interval written as $0/1$ and $1/1$ and then recursively bisecting intervals by means of “Farey mediants”²⁵ $p/q \oplus p'/q' = (p+p')/(q+q')$. The Farey tree rationals can be generated by backward iterates of the number $1/2$ by the Farey map, i.e., starting at $x_0^{(0)} = 1/2$ we get the next layer $x_0^{(1)} = 1/3$ and $x_1^{(0)} = 2/3$ as the two numbers which are mapped to $1/2$ under the Farey map \mathcal{F} . Generally, the 2^n n th inverses of $1/2$ (under the Farey map) are precisely the n th layer of the Farey tree.

Iterating downwards the layers of the Farey tree is to follow the evolution (backwards) in the parameter u and corresponds to one bounce against one of the three walls. Compare with the displayed trees of symbols in M.H. Bugalho *et al* (1986).

A non-chaotic, segment of the type $u \rightarrow u-1$ (a repeated bounce against two given walls) corresponds to continue along the given branch of the Farey tree. For example, drifting along the most left branch of the Farey tree corresponds to going backwards through the numbers $x = 1/2 \rightarrow 1/3 \rightarrow 1/4 \rightarrow 1/5 \rightarrow \dots$ which in the forward direction translates to the segment of evolution $\rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2$ in the u parameter.

The chaotic part (an inversion $u \rightarrow 1/u$) is associated with a shift in “direction” of motion down the tree, that is when we *zig-zag* our way down the tree. For example, the $[1, 1, 1, 1] = [1]^\infty$ orbit, which only contains one oscillation in every major cycle, would here be described as an infinite zig-zag route (each zig and each zag consist of one step, interchanging without end) and the associated Farey number then converges to the golden mean $u = (\sqrt{5} - 1)/2$. This is the initial starting value of the parameter u which in the forward direction generates this orbit.

Each branch of the Farey tree is *similar* to the entire tree. In the gravitational

²⁵See p. 23 in Hardy and Wright (1938) and Cvitanovic and Myrheim (1989).

collapse dynamics governed by the *classical* vacuum Einstein equations there is also no scale (the vacuum Einstein equations are scale invariant²⁶). When the evolution starts to go into the described sequence of oscillations, it does so forever and does not distinguish the situations where 10 or 10.000 major cycles have passed. According to *classical* general relativity nothing prevents this sequence of (self-similar) oscillations from going on “forever”.²⁷

The gravitational collapse (3) thus “organizes itself” (is attracted) into the described (self-similar) never ending sequence of oscillations of the scale functions without any strong fine-tuning of initial-conditions (as regards the amount of matter²⁸ in the model or the initial “set up” of the gravitational degrees of freedom).

It has been stated by Belinskii *et al* that the (self-similar) oscillatory character is a generic “attractor” (locally, i.e. in the neighborhood of every space point) even in the larger space of spatially inhomogeneous gravitational collapses. (See e.g. Belinskii *et al* (1982) and Zel'dovich and Novikov (1983) §23.3). This claim is, however, a subject of substantial controversy.

2.7. Stability properties of the metric (evolved towards and away from the spacetime singularity). Does it isotropize?

It was demonstrated some time ago, E.M. Lifshitz (1946), by means of linear perturbation analysis, that the FRW cosmological solution is indeed *stable* against *local* perturbations of the metric (perturbations over small regions of the space, i.e. over regions whose linear dimensions are small compared to the FRW scale factor R) in the *forward* time direction (away from the singularity), but *unstable* in the direction towards the space time singularity!

In the “class of perturbations”, studied here, which one may call *global* anisotropic $SU(2)$ homogeneous perturbations of the compact FRW solution, the study shows that in the direction towards the singularity a small anisotropy will grow up²⁹ and the gravitational collapse will go into the described BKL-sequence of oscillations, which is *unpredictable*, even if we knew the initial data with an (almost) infinite amount of precision!

In the direction *away from the singularity* the matter terms get increasingly important and the mixmaster cosmological model *isotropizes*³⁰ (cf., e.g., MTW §30.3-30.5,

²⁶There is no natural length associated with general relativity (G, c) and no natural length associated with the quantum principle (\hbar), individually. The union of general relativity and “the quantum” (G, c, \hbar) does contain a natural length, $L_{Pl} = \sqrt{(\hbar G)/c^3}$. (The Planck length).

²⁷However, if we put an initial scale of the gravitational collapse in “by hand” (e.g., for instance, $L = c/H \sim 10^{28}$ cm which is a characteristic scale in the present universe) a restriction concerning the validity of the classical field equations themselves may reduce the number of such oscillations to be very small before Planck scales are reached and quantum fluctuations of the metric become important.

²⁸The perfect fluid matter contributions (in the case of dust or a radiation fluid) in the toy-model gravitational collapse become (in the direction towards the space-time singularity) unimportant at some point (the collapse is curvature dominated) and the model evolves into the scale-free oscillatory (chaotic) collapse dynamics.

²⁹This also happens in the well known “Taub-solution” which is an axisymmetric special case ($a = b$, $c \neq a \& C.$) of the *vacuum* mixmaster metric. This Taub solution is integrable (Taub, 1951) and unstable!

³⁰By “isotropization” we here mean that the anisotropic spacetime metric (3), coupled to perfect fluid matter, evolves into a stage of (quasi) isotropic expansion rate, that is, the metric (3) in the forward time direction evolves into a stage where the expansion is approximately uniform in all directions and is described by Hubble’s law. Note, however, that the curvature of the three-dimensional space is very

Ya.B. Zel'dovich and I.D. Novikov (1983), chapt. 22.7, and V.N. Lukash (1983)) - though not fast enough to explain the remarkable degree of isotropy we see in the Universe today! An idea like, e.g., a Guth/Linde *inflationary* phase of the universe model is needed. However, while at the one hand inflation occurs in anisotropic cosmological models under a wide variety of circumstances, there are actually some difficulties in inflating the compact mixmaster model universe, if the initial anisotropy is too large, cf., e.g., sec. 6.2 in the review by K.A. Olive (1990).

3. CHAOTIC ASPECTS OF THE GRAVITATIONAL COLLAPSE (POSITIVE LYAPUNOV EXPONENTS AND ALL THAT)

Can we assign an invariant meaning to chaos in the general relativistic context? It will be clear in the following that this program of research is still in its infancy and to pursue this question will require a good deal of technical apparatus in general relativity.

A standard probe of chaos for dynamical systems with few degrees of freedom is to look at the spectrum of Lyapunov exponents, in particular the principal (maximal) Lyapunov exponent defined in the following way (J.P. Eckmann and D. Ruelle (1985)): Given a flow $f^t : \mathcal{M} \rightarrow \mathcal{M}$ on a manifold \mathcal{M} and a metric (a norm) $\|\cdot\|$ on the tangent space $T\mathcal{M}$, we define for $x \in \mathcal{M}, \delta x \in T_x\mathcal{M}$,

$$\lambda(x, \delta x) = \lim_{t \rightarrow \infty} \frac{1}{t} \log \|(D_x f^t) \delta x\| .$$

For any ergodic measure ρ on \mathcal{M} this quantity for ρ -almost all $x \in \mathcal{M}$ and almost all $\delta x \in T_x\mathcal{M}$ defines the principal (maximal) Lyapunov exponent for the flow f^t w.r.t. the ergodic measure ρ (and is independent of x and δx).

Note, that the maximal Lyapunov exponent, and more generally the spectrum of Lyapunov exponents, extracted from non-relativistic *Hamiltonian* flows is *invariant* under non-singular *canonical* coordinate transformations which do not involve transformations of the time coordinate (see also e.g. H.-D. Meyer (1986)).

In the general relativistic context we may identify some obstacles for this construction of a “Lyapunov exponent” (e.g. for the mixmaster gravitational collapse, the orbits being three-metrics, ${}^{(3)}g_{ij}$, evolving in “time” towards the “final crunch”):

- (1) *What should we choose as a distance measure “ $\|\cdot\|$ ” on the solution space?*
- (2) *What should we choose as a time parameter?*
- (3) *Is there a natural ergodic measure on the solution space?*

At first, it is natural to try to treat the evolutionary equations (7) for the mixmaster metric (3) as a set of ordinary differential equations on *equal footing* with other dynamical systems governed by some set of ordinary differential equations and apply the standard probes of “chaos” available, in particular the standard methods for extracting Lyapunov exponents.

We decompose the equations (7) into the form of first order differential equations, $\vec{x}' = \vec{f}(\vec{x})$, $\vec{x} \in \mathbf{R}^6$. Write, e.g., the vacuum equations (7) as $2\alpha'' = (e^{2\beta} - e^{2\gamma})^2 - e^{4\alpha}$ (and cyclic permutations) and get for the 6-dimensional state vector

$$\vec{x} = \vec{x}(\tau) = (\alpha, \beta, \gamma, \alpha', \beta', \gamma') \tag{24}$$

anisotropic and, as a rule, does not isotropize. See also Zel'dovich and Novikov, §22.7.

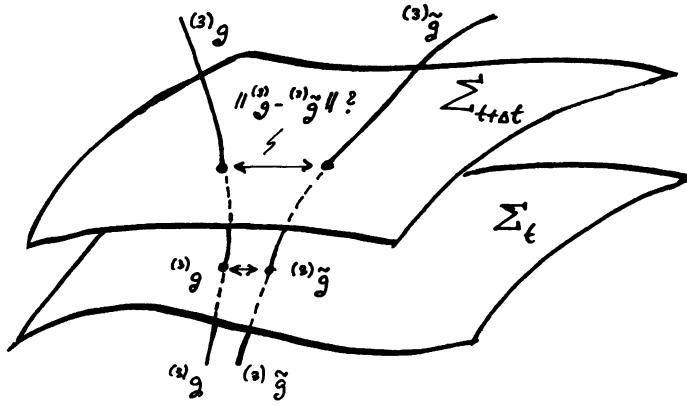


Figure 7. What is a *natural* distance measure $\|{}^{(3)}g - {}^{(3)}\tilde{g}\|$ between two nearby space-time metrics ${}^{(3)}g$ and ${}^{(3)}\tilde{g} = {}^{(3)}g + \delta {}^{(3)}g$, where g and \tilde{g} are both solutions to the vacuum Einstein equations but correspond to slightly different initial conditions, say? In the case of non-relativistic Hamiltonian dynamical systems, measures of “chaos” such as principal Lyapunov exponents are calculated by using an *Euclidean* distance measure which is naturally induced from the structure of the kinetic energy term in the non-relativistic Hamiltonian. In the general relativistic context, however, it is not obvious why one should use an Euclidean distance measure to assign a distance between two three-metrics. Do we need such a distance measure on the space of mixmaster collapses to talk about chaos?

the coupled *first order* equations

$$\frac{d}{d\tau} \begin{pmatrix} \alpha \\ \beta \\ \gamma \\ \alpha' \\ \beta' \\ \gamma' \end{pmatrix} = \begin{pmatrix} \alpha' \\ \beta' \\ \gamma' \\ \frac{1}{2}(e^{2\beta} - e^{2\gamma})^2 - \frac{1}{2}e^{4\alpha} \\ \frac{1}{2}(e^{2\gamma} - e^{2\alpha})^2 - \frac{1}{2}e^{4\beta} \\ \frac{1}{2}(e^{2\alpha} - e^{2\beta})^2 - \frac{1}{2}e^{4\gamma} \end{pmatrix} \quad (25)$$

which are supplemented with the first integral constraint $I = 0$ with I given in (11) as a constraint on the initial state vector for the gravitational collapse.

Calculations of Lyapunov exponents were in all previous studies (cf., e.g., Burd et al. (1990), Hobill *et al* (1991), Berger (1991), Rugh (1990a), see also Rugh and Jones (1990)) based on the use of an Euclidean distance measure of the form,

$$\|\vec{x}(\tau) - \vec{x}^*(\tau)\| = \sqrt{\sum_{i=1..6} (x_i(\tau) - x_i^*(\tau))^2} \quad (26)$$

assigning a distance between the two state vectors $\vec{x}(\tau), \vec{x}^*(\tau) \in \mathbf{R}^6$ in the phase space.³¹

This choice of Euclidian distance measure is directly inspired from the study of ordinary (non-relativistic) Hamiltonian systems. However, in general relativity there is a priori no reason why one should use such a distance measure to give out the distance between ${}^{(3)}g$ and a “nearby” three-metric ${}^{(3)}g + \delta {}^{(3)}g$. A distance measure like (26) is gauge dependent, i.e. *not invariant* under a change of coordinates, $g \rightarrow \tilde{g}(g)$. (Should

³¹What is equivalent Lyapunov exponents were extracted from calculating the Jacobian matrix from the flow equations (25) of the state vector $(\alpha, \beta, \gamma, \alpha', \beta', \gamma')$ and integrating up along the orbit. Cf. e.g. appendix A in S.E.Rugh (1990a).

the distance between two different “mixmaster collapses” (at some given time) be an artifact of the chosen set of coordinates for the description of the collapse?).³²

If the maximal Lyapunov exponent is extracted from the flow equations (25) for the mixmaster collapse with a Euclidian distance measure (26) on the solution space we obtain the following result (from S.E. Rugh (1990a)):

In τ -time: With respect to the time parameter $\tau = \int dt/abc \sim \ln t$ introduced as a standard time variable (cf. Landau and Lifshitz (1975) §118) for the description, the approach to the singular point is *stretched* out to infinity ($\tau \rightarrow \infty$). Despite the mentioned aspect of chaotic unpredictability of the model, as described by the Farey³³ and Gauss map connected to the combinatorial model (22) for the axes, this *stretching* is “so effective” as to make the Lyapunov characteristic exponent, extracted from the $(\alpha, \beta, \gamma, \alpha', \beta', \gamma')$ phase flow, zero in τ -time: $\lambda_\tau = 0$. (See S.E. Rugh (1990a), S.E. Rugh and B.J.T. Jones (1990), D. Hobill *et al* (1991)³⁴ and also A. Burd, N. Buric and R.K. Tavakol (1991)). This is not standard for a chaotic deterministic model, but it appears that the zero Lyapunov exponent is simply a consequence of the time variable chosen for the description.

In t -time: In the original synchronous time parameter $t = \int abc d\tau$ the scale functions of the metric

$$ds^2 = -dt^2 + \gamma_{ab}(t) \omega^a(x) \omega^b(x) , \quad \gamma_{ab}(t) = \text{diag}(a^2(t), b^2(t), c^2(t))$$

exhibits an infinite sequence of oscillations and the successive major cycles condense infinitely towards the singularity which we, without loss of generality, may take to be at $t = 0$. (See also the very illustrative picture in Kip S. Thorne (1985)) of oscillating, fast growing, tidal field components of the successive oscillations as the spacetime metric approaches the singular point at $t = 0$).

In any finite time interval $t \in [T, 0]$, where T denotes an arbitrary small positive number, the model exhibits an infinite, unpredictable sequence of oscillations.

A Lyapunov characteristic exponent, although ill defined, is unbounded in t time. (A perturbation will be amplified faster than exponentially with respect to the metric t time and with respect to the distance measure (26) on approach to the singularity).

The quoted Lyapunov exponents above are for mixmaster collapses governed by the *vacuum* Einstein equations corresponding to the vanishing $I = 0$ of the first integral (11) or corresponding to mixmaster collapses coupled to perfect fluid matter with $p \leq \frac{2}{3}\rho$. For initial conditions which fail to satisfy the first integral

³²Being somewhat acquainted with general relativity and its Hamiltonian formulation, a more natural choice may well be a distance measure which is induced from the structure of the Hamiltonian in general relativity (just like the Euclidian distance measure (26) is induced from the Hamiltonian (the kinetic term) in ordinary *non-relativistic* Hamiltonian dynamics), i.e. a distance measure of the type $ds^2 = \int d^3x G_{ijkl} \delta g^{ij} \delta g^{kl}$ where G_{ijkl} is the Wheeler superspace metric on the space of three metrics. However, such a distance measure is not positive definite. (Sec.3.8).

³³We have seen that the Farey map comes closer to mimic the bounce dynamics in terms of a simple, non-invertible one-dimensional map and it has $\lambda = 0$. Moreover, the $\tau = \int dt/abc$ time between each bounce grows very fast, so it is no surprise to find $\lambda_\tau(\alpha, \beta, \gamma, \dots) = 0$. Moreover, since reliable estimates of the time intervals between bounces may be found in the literature (cf., e.g., Khalatnikov *et al.* (1985) and references therein) it is possible to show this result by upper bound estimates.

³⁴Note, that Dave Hobill *et al* have extracted the entire spectrum of all six Lyapunov exponents from the phase flow - and not only the principal value (the maximal Lyapunov exponent).

constraint $I = 0$, the character of the solutions depends on the sign of the first integral I . We have qualitatively different behavior for $I < 0$, for $I > 0$ and for $I = 0$. For $I < 0$ (inclusion of negative “stiff matter” energy densities) the complicated behaviors in fig.1-3 are reflected in positive Lyapunov exponents, with respect to the τ time parameter, extracted from such numerical solutions. From the scale invariance of the Einstein equations, it follows that if one “unwittingly” put an amount $p = \rho = \rho_0 = (8\pi G)^{-1}I < 0$ of negative mass-density in the model, the maximal Lyapunov exponent roughly scales as

$$\lambda \sim \sqrt{-\rho_0} .$$

This is exactly what G. Francisco and G.E.A. Matsas (1988) “unwittingly” have plotted as fig.4. in their paper.

For $I > 0$ (i.e. the inclusion of “stiff matter” with positive energy density) all three scale functions will decrease monotonically after some transient τ time and the maximal Lyapunov exponent will be zero. The reason why even V.A. Belinskii and I.M. Khalatnikov (1969) had a minor accident in obtaining qualitative accordance between numerical simulations and their theoretical model is due to the fact that they had an extremely small error in choosing the initial data. They had $I = +5.5 \times 10^{-2} > 0$, which prevents the spacetime metric from going into the described sequence of oscillations according to the *BKL-combinatorial* model.

For the values of the Lyapunov exponents associated with the vacuum gravitational collapse one notes that the fact that Lyapunov exponents, as indicators of chaos, are not invariant under time reparametrizations is no surprise:

A Lyapunov exponent (which is a “per time” measure of having exponential separation of nearby trajectories in “time”) has never been invariant under transformations of the “time” coordinate!

Moreover, for a given fixed choice of time variable, transformations of the *coordinates for the three-metric g_{ij}* on the spacelike hypersurfaces Σ_t should be accompanied by a transformation of the distance measure $\|\cdot\|$ on the three-space. If one does not take proper account of transforming the distance measure (the metric) on the solution space when performing a *coordinate transformation* to new variables (e.g. transformation from the ADM-variables to the Chitre-Misner variables on the Poincaré disc), then some definite value of a Lyapunov exponent (e.g. $\lambda_\Omega = 0$ in the ADM-variables) may be transformed into any other value by that coordinate transformation (and such measures of “chaos” which are artifacts of the coordinate transformations are not strong candidates for capturing anything interesting about inherent properties of a dynamical system.³⁵

³⁵Thus, when J. Pullin (1990) arrived at the conclusion that the mixmaster toy-model gravitational collapse is chaotic because a positive Lyapunov exponent “ $\lambda = 1$ ” can be associated with the geodesic motion on the Poincaré disc one should be careful to project out (separate) what are artifacts of the coordinate-transformations and what are inherent properties of the dynamical system. (Invariant under gauge choices). Certainly the element of chaos and unpredictability of the mixmaster collapse is not (alone) due to the negatively curved interior of the Poincaré disc. Even the completely integrable Kasner metric may be mapped to the interior of the negative curved Poincaré disc (by a set of hyperbolic coordinate transformations identical to those applied to the mixmaster metric). No chaos arises, since there is no scattering in any potential boundaries. However, as an artifact of not transforming the distance measures under the coordinate transformation, we suddenly appear to have local exponential instability in all directions on the Poincaré disc - despite that the Kasner metric is completely integrable with linearly growing separation in the original ADM-variables.

3.1. Chaos in the “coordinates” or in the gravitational field?

The fact that Lyapunov-exponents are found to be strongly gauge dependent “touches” a deeper problem (S.E. Rugh (1990 a,b) connected with the characterization of “chaos”, or other dynamical characteristics, in theories like general relativity (and to some extent in gauge-theories) - not present in, e.g., hydrodynamical turbulence studies: How can you be sure that your measure of “chaos” is not some artifact of the “gauge” variables chosen? That is: Does the apparent “chaos” originate from the gauge-variables chosen (“chaos in the semantics”) or is it present in the “real world”?

To distinguish (in a coordinate invariant way) whether the “metrical chaos” is due to a “chaotic choice of coordinates” (we may call it “chaos in the semantics”) or reflects “real chaos and turbulence” in the gravitational field does not seem to be an easy task, as this separation between coordinate choices and gravitational fields is indeed very tricky business in general.

This question is, in some sense, the “chaos in general relativity” (“metrical chaos”) analogue of Eddingtons worries about the “spurious gravitational waves” as a propagation of “coordinate changes” with the “speed of thought”.

It is not difficult to invent (apparently) “chaotic solutions” of “space-time metrics” where the “chaotic signal” is a pure artifact of gauge (one may consider, e.g., a 2+1 dimensional study of pure “gravity” in a non-linear gauge). H.B. Nielsen and S.E. Rugh (1992 a).

The problem is rooted in the more general question (which is itself very interesting):

How gauge-invariantly may we capture any dynamics of the gravitational field (chaotic or not) in the Einsteins theory of general relativity?

It is well known that it is very difficult to extract gauge-invariant quantities *locally* from general relativity - even in the weak field limit!³⁶

While the collection of the (non-local) Wilson loop variables

$$Tr(P \exp \oint_{\gamma} g A_{\mu}^a(x) \frac{\lambda^a}{2} dx^{\mu})$$

in principle offer a tool to characterize the dynamics of non-Abelian gauge theories in a complete gauge invariant manner,³⁷ such a “projection out” of “the gauge-invariant content” is troublesome in gravity. Like in Yang-Mills theories, we may attempt to characterize the gravitational field invariantly by loops (capturing the curvature tensor components $R_{\beta\gamma\delta}^a$'s, just as the loops in Yang-Mills theories capture the $F_{\mu\nu}$'s). But there is a problem of saying (in a diffeomorphism invariant way) *where* the loop is!

One can construct completely gauge-invariant and *global* quantities like for example

$$I(\eta) = \int d^4x \sqrt{-g} \delta(C_{\alpha\beta\gamma\delta} C^{\alpha\beta\gamma\delta} - \eta) \quad (27)$$

³⁶For example, Roger Penrose has for several years attempted to extract a measure of “entropy” in the metric field itself. But it is not easy to make a *gauge-invariant* and sensible construction. So far, Roger Penrose has not succeeded. Pers. comm. with Roger Penrose (at the Niels Bohr Inst.).

³⁷Thereby loosing locality, cf. also the Aharonov-Bohm effect.

which measures the invariant 4-volume which has the value $C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta} = \eta$. But it is very hard to capture any dynamics with such a (space-time global) signal.

One may question the very suitability (in more generic cases) of concepts like

- A Lyapunov exponent “ λ ” (a “per time” measure characterizing the “temporal chaos”: does the “distance” between “nearby” orbits grow up exponentially with time?)
- A spatial correlation length “ ξ ” (in a spatially inhomogeneous gravitational field we may have, that domains separated by distances considerably larger than $\sim \xi$ cannot “communicate” due to spatial disorder in the field configurations: “metric turbulence” indicates a lack of correlations between gravitational signals in time as well as in space.⁴ This “spatially disordered” aspect of chaos falls away for the spatially homogeneous toy-models).

To the extent that we may split the spacetime metric $(^4)g$ into a three-metric $(^3)g$ evolving in a “time”⁵ the very concept of, e.g., a Lyapunov exponent may perhaps be applicable, though *gauge-invariant generalizations* (i.e. measures of “metrical chaos” invariant under general space-time coordinate transformations or some *more restricted* class of coordinate transformations) are (obviously) better to have - if they are possible to construct?

It may very well be that one should head for a “no go” theorem for the most ambitious task of constructing a generalization of a Lyapunov exponent (extracted from the continuous evolution equations) which is meaningful and invariant under the full class of spacetime diffeomorphisms (H.B. Nielsen and S.E. Rugh). One should therefore, at first, try to modify the intentions and construct (instability) measures which are invariant under a smaller class of diffeomorphisms (e.g. coordinate transformations on the spacelike hypersurfaces Σ_t which do not involve transformations of the time coordinate).

In the mixmaster toy-model context of a spatially homogeneous spacetime metric this question is particularly simple to address since all the physical signals are *functions of time* only.

The need for gauge invariant measures of chaos in this context of general relativity was noted and discussed in S.E. Rugh (1990a), emphasized as a major point in S.E. Rugh (1990b) and, also, emphasized in J. Pullin (1990). See also discussions in e.g. H.B. Nielsen and S.E. Rugh (1992), in M. Biesiada and S.E. Rugh (1994) and in S.E. Rugh (1994).

1.2. Chaos in ever expanding (or ever collapsing) phase spaces

In the description of the gravitational collapse involving the $(\alpha, \beta, \gamma, \dots)$ phase flows (the parametrization in, e.g., Landau and Lifshitz (1975)) and the $(\Omega, \beta_+, \beta_-, \dots)$ phase

⁴Often (cf., e.g., T. Bohr (1990)) temporal and spatial chaos are somewhat connected in a relation of the type: $\xi \sim c/\lambda$ where c is the velocity of propagation (velocity of “light” in the case of propagation of disturbances of spacetime curvature in general relativity).

⁵Note that we have been able to study our simple spatially homogeneous toy-model gravitational collapse in a (*global*) *synchronous* frame of reference. One can always choose a synchronous reference frame *locally*, but in generic situations, for strong gravitational fields, a global construction of a synchronous frame of reference easily breaks down because of the tendency of these coordinates to focus to a *caustic* (and a coordinate singularity will develop).

flows (the anisotropy-variables of C.W. Misner et al.), there is no “recurrence” of orbits because the phase space is ever expanding⁴⁰, and it is not straightforward to define what is meant by a concept like “ergodicity” or the “mixing property”, for example, in the outward expanding Hamiltonian billiard in the ADM-Hamiltonian description. In this sense the gravitational collapse study presents a dynamical system, which differs from most other systems treated in the theory of non-linear dynamical systems. Thus our gravitational collapse orbit is also a playing ground for investigating how to deal with chaos in expanding (or collapsing) phase spaces.

The ideas of chaos apply most naturally to time evolutions with “eternal return”. These are time evolutions of systems that come back again and again to near the same situations. In other words, if the system is in a certain state at a certain time, it will return arbitrarily near the same state at a later time. (Cf. D. Ruelle (1991), p. 86)

How do we extract measures of chaos and ergodicity in an ever expanding phase space?

- As we have already described, one may project out (extract) lower dimensional signals from the flow which may capture (chaotic) “return properties” of the dynamics. An example is the variable “ $x = 1/u \in [0, 1]$ ” which we extracted from the phase space trajectories in sec.2.4.
- If we characterize the dynamical evolution of the gravitational collapse in terms of the infinite bounce sequence (i.e. encoding the gravitational collapse by an infinite collision sequence against the three wall boundaries) we hereby map an ever expanding phase space (in the $(\alpha, \beta, \gamma, \dots)$ variables, say) to an infinite symbolic sequence with return properties (e.g. with periodic orbits).

3.3. The many gauges and different variables describing the mixmaster gravitational collapse

The amount of work which has been put in the invention of different descriptions (different gauges) and different approaches is substantial: Cf. the study of the mixmaster gravitational collapse in the ADM variables (cf. e.g. B.K. Berger (1990)), the Misner-Chitre description (J. Pullin (1990)), the BKL-variables (D. Hobill *et al* (1991), S.E. Rugh (1990 a)), the Ellis-MacCallum-Wainwright variables (A.B. Burd, N. Buric and G.F.R. Ellis (1990)), Ashtekhar variables (Ashtekhar and Pullin (1989)), the description by O.I. Bogoyavlensky and S.P. Novikov (e.g. used in the approach by M. Biesiada, J. Szczesny and M. Szydłowski).

The “transformation theory of chaos” (which concepts of “chaos” are invariant under the transformations?) between the different gauges and approaches is not easy. As we have seen, in some gauges there is a positive (finite or infinite, depending on the gauge choice) value of a “Lyapunov exponent” - in other gauges it is zero. In many gauges the phase space is *ever expanding* and a notion like that of “ergodicity” (relying by definition on return properties of the flow) can hardly be assigned any meaning at all. In other gauges, e.g. the “Misner-Chitre” gauge, the dynamics of our spacetime metric

⁴⁰If we follow this metric to the final crunch singularity, the configuration space variables (a, b, c) shrink to zero volume ($abc \rightarrow 0$), while if we take the logarithmic scale factors $(\alpha, \beta, \gamma) = (\ln a, \ln b, \ln c)$ the configuration space is not bounded from below (we have $\alpha + \beta + \gamma \rightarrow -\infty$). In terms of anisotropy variables, also commonly used, $(\Omega, \beta_+, \beta_-)$ the anisotropy $\|\vec{\beta}\|$ grows to infinity and the trajectory of anisotropy $\vec{\beta}$ does never return to some given value.

Table 2. Examples of the many descriptions of the very same gravitational collapse. What are the “invariant” properties of the dynamics which can be extracted from all these descriptions?

The many different gauges and approaches
<ul style="list-style-type: none"> • BKL variables (Landau and Lifshitz) • The ADM-Hamiltonian variables (MTW) • The Misner-Chitre coordinates • Ellis-MacCallum-Wainwright variables • Ashtekhar variables • The approach by Bogoyavlensky and Novikov

(3) is, it appears, ergodic and, even, a K-flow! (cf., e.g., J. Pullin (1990)). This rather “messy situation” is, of course, an artifact of not having constructed gauge invariant measures of chaos (or ergodicity) in this context of general relativity.

Yet, in the mixmaster gravitational collapse, the “bounce” dynamics is there and that we must be able to capture in a gauge invariant way (i.e. that ∞ many bounces take place before the spacetime singularity is reached).

Imprints of the bounce structure of the mixmaster dynamics should be seen in various invariants which can be extracted from the mixmaster metric, e.g. in the Weyl curvature invariant $C^2 = C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}(\tau)$ of the mixmaster metric. Thus, in order to illustrate how a “good physical signal” behaves for a typical trajectory we have set out to calculate and display (cf. Biesiada and Rugh (1994)) this Weyl curvature invariant. Premature investigations indicate that if $C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}$ is plotted with respect to the time variable τ , one observes (Biesiada and Rugh (1994)) that the bounce structure is also seen on this curvature invariant.⁴¹

Also other (higher order) invariants could be constructed for the mixmaster gravitational collapse and one should verify whether the physical degrees of freedom of the mixmaster collapse in principle may be captured *exhaustively* in terms of the algebraic invariants constructed from the Weyl curvature tensor field.

3.4. On symbolic dynamics for the bounce structure of the gravitational collapse

Let us now consider a particular “gauge”, say, the ADM Hamiltonian description (or the “Misner Chitre” gauge) in which the gravitational collapse is described as a “ball” exhibiting an infinite sequence of “bounces” against a potential boundary which is outward

⁴¹During the workshop it was discussed (Charles W. Misner, Piotr Chrusciel *et al*) that it is not fully investigated how the Weyl curvature tensor behaves near the generic singularity of the mixmaster collapse. For example, $C^2 = C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}(\tau)$ will not necessarily increase *monotonically* towards the big crunch singularity. The near singularity behavior of the Weyl tensor is therefore interesting in itself and deserves further examination.

expanding⁴² (or stationary in the “Misner Chitre” gauge). Is a “bounce” well defined? Recall, that the scattering walls - which effectively arise due to the three-curvature scalar 3R on the space-like sections of the metric $ds^2 = -dt^2 + \gamma_{ij}(t)\omega^i(x)\omega^j(x)$ - turn into “infinitely” hard walls when sufficiently near the space-time singularity of the metric. Hence, the dynamics resembles that of an ideal billiard and it gets easier and easier to define when the orbit “bounces” against the potential wall as the metric approaches the singular point.

It may be useful to assign a *symbolic dynamics* to the dynamical system, i.e., a scheme that assigns a unique symbolic string (coding) to each orbit. The simplest qualitative way to describe an orbit is to list the order in which it hits the three boundary walls $\partial\mathcal{B}_1, \partial\mathcal{B}_2, \partial\mathcal{B}_3$ using the wall labels as symbols (characterizing the orbit by its infinite collision sequence). This yields a description of the orbit in terms of ternary symbolic dynamics:

“1” for scattering against the boundary $\partial\mathcal{B}_1$ (where $c \gg a, b$)

“2” for scattering against the boundary $\partial\mathcal{B}_2$ (where $a \gg b, c$)

“3” for scattering against the boundary $\partial\mathcal{B}_3$ (where $b \gg a, c$)

As a randomly chosen reference example of a segment of a “symbolic string” (compare with Kip. S. Thorne (1985)) we may, for example, have

$$\dots \underbrace{12121212}_{4 \text{ oscillations}} \quad \underbrace{131313}_{3 \text{ oscillations}} \quad \underbrace{23232323232323}_{8 \text{ oscillations}} \quad \underbrace{121\dots}_{\text{etc}} \dots \quad (28)$$

The sequence of oscillations between two walls leading to one of the three corners is denoted as a “(major) cycle”. Thus, the displayed segment (28) has 4 oscillations in the first of the displayed (major) cycles (number “ i ”, say) 3 oscillations in the next major cycle (number “ $i + 1$ ”), 8 oscillations in cycle “ $i + 2$ ”, etc.

Binary symbolic dynamics suffices

Due to the grammar of the symbolic dynamics, which forbids the appearance of ...11..., ...22... or ...33... in a sequence (no orbit can hit the same wall twice without hitting another wall first) the three-letter alphabet (above) may be reduced to a *two-letter alphabet*. This reduction can be accomplished in many ways. One possible way of assigning binary symbolic dynamics to the dynamics is the following: Assign a “0” for scattering against the boundary $\partial\mathcal{B}_1$ and a “1” for scattering against the boundary $\partial\mathcal{B}_2$. Scattering against the last of the three boundaries $\partial\mathcal{B}_3$ we denote “0” if it follows a bounce against the boundary $\partial\mathcal{B}_1$ and denote “1” if following a bounce against the boundary $\partial\mathcal{B}_2$. In this encoding the motion is considered as a series of choices of what boundary to go to from the present boundary (there are always two possibilities, and binary symbolic dynamics therefore suffices to describe the chosen route). Except for some arbitrariness in the beginning of a symbol sequence, this gives a one-to-one correspondence between the *pruned*⁴³ ternary dynamics and an unrestricted binary dynamics.⁴⁴

⁴²Cf., e.g., MTW §30.7 or Ryan and Shepley (1975). A detailed description of the ADM-Hamiltonian description of the gravitational collapse is also given in S.E. Rugh (1990a) (which is available upon request).

⁴³*Pruning*: If some symbolic sequences have no physical realizations, the symbolic dynamics is said to be “pruned”. In that case the symbolic alphabet must be supplemented with a set of grammatical rules (i.e. a set of pruning rules). If all possible symbol sequences correspond to physical trajectories, the symbolic dynamics is said to be “complete”. In a systematic encoding process, one will ask first if

For instance, our reference-segment (28) reads

$$\dots \underbrace{01010101}_{4 \text{ oscillations}} \underbrace{000000}_{3 \text{ oscillations}} \underbrace{1111111111111111}_{8 \text{ oscillations}} \underbrace{010\dots\dots}_{\text{etc}} \quad (29)$$

If we were to carry out a systematic exploration of system characteristics in terms of the periodic orbit structure (this includes finding periodic orbits up to some given length, quantize the system in terms of the periodic orbits, Selberg zeta functions and all that) it would be of importance to use the minimal symbolic dynamics (with no pruning). We shall, however, just use the direct symbol assignment **1**, **2**, **3** in order to have a fixed notation.

3.5. Chaos and the topological structure of the gravitational collapse

A simple example illustrates that a zero Lyapunov exponent is not contradictory to an intrinsic unpredictable and chaotic behavior (from S.E. Rugh (1990 a)).

Consider, for example the famous Lorentz attractor generated as an asymptotic attractor by the Lorentz time evolution equations (Lorentz, 1963) - just like the (BKL) oscillatory behavior is generated as an asymptotic attractor⁴⁵ of the vacuum Einstein equations for the mixmaster spacetime metric (3).

The “particle” performs an infinite number of consecutive rotations about two rotation centers, schematically shown in fig.8. It rotates, lets say, n_1 times around the left rotation center (“first major cycle”) and then n_2 times around the right rotation center (“second major cycle”) and return to the left again ...etc...etc...ad infinitum.

One feature of the non-linear Lorentz system is that the number of consecutive rotations about the two rotation centers is unpredictable, since these numbers (n_1, n_2, \dots) depend sensitively on initial conditions. (See also Lichtenberg and Lieberman (1983), p.59-62). In that sense the behavior of the system is *intrinsically* chaotic (stochastic) and not predictable. With respect to the standard time parameter of the Lorentz evolution equations, a maximal Lyapunov characteristic exponent, extracted from the flow, mirrors this unpredictability and is greater than zero.

However, by making some kind of an “exponential stretching” of the time parameter axis (thereby making the particle moving around slower and slower with respect to the new time parameter) it is easy to make the corresponding Lyapunov exponent, extracted from the new first order autonomous differential flow, zero. The unpredictability of the number of consecutive rotations performed in each “major cycle” of course still remains!

The gravitational collapse, in the oscillatory BKL-regime on approach to the singularity, exhibits consecutive rotations around 3 “centers” (there are three “attraction centers”: α and β can oscillate (γ declines), α and γ can oscillate (β declines) or β and

there is *at most* one orbit to a given symbolic sequence. If this is the case we have a *covering* symbolic dynamics. If there *is* also one orbit for any symbolic sequence, the symbolic dynamics is said to be *complete*.

⁴⁴Note, that in the three disc problem on the hyperbolic Poincaré disc we have “full symbolic dynamics” while in the three-disc problem on an *Euclidian* space (i.e. if the billiard had been in Euclidian space) there exist binary symbolic sequences which do not correspond to physical trajectories. In that sense, the hyperbolic three-disc dynamics is simpler to work with than the Euclidian three-disc problem. See also Giannoni and Ullmo (1990).

⁴⁵The “basin of attraction” are initial conditions which are not axisymmetric (cf. the Taub-solution).

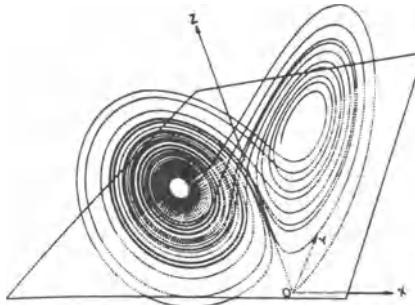


Figure 8. The intrinsic unpredictability of the number of consecutive rotations about the two rotation centers is *independent* of the choice of time variable (whereas the spectrum of Lyapunov exponents is not).

γ can oscillate (α declines)) corresponding to oscillations in one of the three channels of the billiard game in the Hamiltonian description. (See, also, fig.9 of the triangular symmetric scattering domain in the Poincaré disc). If we, by the number of rotations around a “center”, mean the number of oscillations in a period (a major cycle) where two scale functions oscillate and the third declines, the BKL-model will be intrinsically unpredictable as the Lorentz attractor described above (the number of consecutive rotations around now three “rotation centers” depends sensitively on initial conditions, and the “per cycle” entropy is very big: $h = \pi^2/6 \ln 2 \gg 0$). But the “particle” moves too slow with respect to the standard time parameter (constructed from the metric t time) $\tau = \int dt/abc$, as $\tau \rightarrow \pm\infty$, to mirror this unpredictability.

The property of a dynamical system to behave “chaotically” ⁴⁶ should be an intrinsic property of the system and not rely on the variables chosen for the description.

3.6. Which concepts should characterize “chaos” in the mixmaster toy-model collapse (and in more general cases)?

There is nothing mysterious in selecting a particular “gauge” and describing physics, e.g. gravitational collapse physics, in that particular gauge. Thus, very often, it is not very useful to do the calculations in a gauge-invariant way (and in general relativity it may in fact not be possible). So one, usually, “fixes the gauge” and performs the calculations in that gauge.

However, in my opinion, it is of interest to pursue some routes in order to study how we may eventually deal more invariantly with the “chaotic” mixmaster gravitational collapse:

- 1.** How little structure on the solution space is needed to deduce information about chaos?
- 2.** One may argue that there is a preferred superspace metric and a preferred supertime (Wheeler’s superspace) as a natural structure on the solution space
- 3.** One may attempt a (more) invariant description (involving

⁴⁶I.e. the system displays a complex evolutionary pattern rather than a simple, integrable pattern (which is “just” quasi-periodic motion in the action-angle variables).

knowledge about extrinsic curvature invariants “ $\text{Tr } \mathbf{K}$ ” etc.).

- 4.** Study our free observer and see what happens with him (tidal stretches etc.) in his local time. That is, after all, a very *physical* question!

I shall here attempt at preliminary sketches of how to pursue the first two routes. We have already, in sec.2.5, sketched the short, yet exciting, story of the free falling astronomer which reaches the spacetime singularity in finite proper time (the tidal stresses exerted on him by falling freely in the collapsing mixmaster spacetime metric should be properly analyzed in terms of the Riemann tensor $R_{\alpha\beta\gamma\delta}$ for the mixmaster metric and not in terms of the Riemann tensor of the Kasner metric).

We shall not pursue route no.3 here but note that in the “3+1” splitting of the spacetime into a three-metric ${}^{(3)}g$ evolving in a “time”, a complete description of the geometry of the four manifold also needs, besides the intrinsic 3-geometry (the induced three-metric ${}^{(3)}g_{ij}$ on the spacelike hypersurfaces Σ_t), knowledge about how the 3-geometry is embedded in the enveloping 4-geometry, which is characterized by the *extrinsic* curvature K_{ij} (cf., e.g., MTW §21.5). One may attempt a more invariant description (i.e. measures of local instability of collapse orbits) by extracting the information from both the intrinsic geometry and the extrinsic curvature and its associated invariants (like its trace, “ $\text{Tr } \mathbf{K}$ ”).

3.7. Is it possible to characterize the complexity of the gravitational collapse without use of a distance measure $\|\cdot\|$ on the solution space? (and without reference to a particular time variable?)

Since the standard measure, the Lyapunov exponent λ , is a “per time” measure, it depends, obviously, on the choice of time parameter ($\lambda = 0$ in $\tau = \int dt/abc$ time while $\lambda = \infty$ in metric t time). Besides, the definition of a Lyapunov exponent requires that we have at hand a metric $\|\cdot\|$ on the solution space, which measures “distances” between (neighboring) trajectories. In the chain of adding more and more structure to the solution space,

$$\left(\begin{array}{c} \text{Topological} \\ \text{Space} \end{array} \right) \subset \dots \subset \left(\begin{array}{c} \text{Manifold Structure} \\ \text{With a Metric} \end{array} \right)$$

one also hereby gradually introduces more and more arbitrariness. We need analysis to which extent one may construct measures of chaos where we do not have to rely on the (arbitrarily) chosen time-parameter “ t ” or some (arbitrary) chosen distance measure “ $\|\cdot\|$ ” on the phase space (cf., also, H.B. Nielsen and S.E. Rugh (1992 a)).

How much “structure” on the “solution space” do we need to get information about whether a collapse dynamics is of a complex (chaotic) or regular (integrable) type, i.e. extract interesting information about “chaos”? Do we really need a manifold structure with a metric $\|\cdot\|$ on it - or can we, from knowledge of the topological structure alone, say, (on the space of solutions) tell something about chaos of interest?

For our gravitational collapse (3) - being spatially homogeneous - all physics is entirely given in terms of *functions of time* (e.g. encoded in “signals” like $C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}(\tau)$ and higher order algebraic invariants) and the phase space dynamics is a low dimensional signal in “time”.

It is natural to try to assign a “good symbolic dynamics”⁴⁷ from the phase space dynamics. I.e. one may think of the dynamics “described” by a string of symbols

$$\{ \dots, S_i, S_{i+1}, \dots, S_{i+n}, \dots \} , \quad S_i \in \{ 1, 2, \dots, n \} . \quad (30)$$

We have already noticed that there is a naturally ternary (or binary without pruning) symbolic encoding of the *bounce structure* of the gravitational collapse. (Thereby, one maps the ever expanding or collapsing phase space (without return properties of the collapse orbit), cf. sec.3.1, to symbol sequences *with* return properties (e.g. having periodic orbits)).

We have a *gauge-invariant* “before” and “after” - but the size of the step in “time” between two symbols in (30) is merely reflecting an arbitrary gauge choice of a “time coordinate”.⁴⁸

Now, let us say that we are given the set of *all possible sequences* of finite length, which may be realized by the physical system (that is usually many!). Can we then deduce, just by looking at these sequences, whether the underlying physical system is integrable or not (i.e. without having any metric $\| \cdot \|$ on the original solution space)?

Complexity measures for symbol sequences are already at hand. The *topological entropy* h_{top} of a string of symbols states how the number $N(n)$ of all the possible sequences of length n grows exponentially with n : $N(n) = e^{h_{top} n}$. It is 0 for an integrable system because of the quasi- or almost-periodicity⁴⁹ (setting severe restrictions on the number of possibilities for strings of length n). For a random string of symbols the topological entropy is equal to the logarithm of the number of symbols (e.g. $\log 2$ for the Bernoulli shift). For the mixmaster gravitational collapse the number of realized symbol sequences (characterized by its infinite bounce sequence) grows with the length n of the symbol string like $\sim 2^n$ and the topological entropy is thus $\log 2 > 0$. That is, for the encoding of the gravitational collapse with the binary (or ternary) symbolic dynamics, e.g. our segment

$$\dots \underbrace{01010101}_{4 \text{ oscillations}} \underbrace{000000}_{3 \text{ oscillations}} \underbrace{1111111111111111}_{8 \text{ oscillations}} \underbrace{010\dots}_{\text{etc}} \dots$$

⁴⁷This is a very difficult point. A priori, the assignment of a symbol-sequence from phase space dynamics (coding the trajectory by a discrete “alphabet”) is arbitrary and accomplished by an arbitrary partitioning of the phase space into different domains (which are assigned different symbols). In many systems it is not easy to find a good partition to encode a generic orbit.

⁴⁸The continuous flow of events during the gravitational collapse is, from the point of view of “causality” (in the sense of which event is before the other) invariant under time reparametrizations and diagnostic tools such as a Fourier transform (an investigation in this spirit is also J. Demaret and Y. De Rop (1993)) of a signal - which can also be performed on symbol sequences - do not need a distance measure $\| \cdot \|$ on the solution space. However, it needs some sort of time variable. Note, that if we take the infinite string of symbols (30) to mean a symbolic encoding of the infinite *bounce* sequence of the mixmaster gravitational collapse we have, hereby, not escaped the introduction of a “time” variable: It is a time variable i (the label of the symbols S_i) which counts one unit for each “bounce” (a recognizable “cosmological event”). It is like “heart beats” of the mixmaster gravitational collapse. In terms of these “heart beats” the gravitational collapse is “infinitely old” when the “big crunch” singularity is reached. See also C.W. Misner (1969 c) and MTW, §30.7, p. 813-814.

⁴⁹Quasi-periodicity, or almost-periodicity, for all realized symbol sequences, is a sign of integrability. Quasi-periodicity (multiply periodic motion) means that the Fourier transform is entirely composed of δ -functions. The finite sums (trigonometric polynomials) $s(t) = \sum_{n=1}^N \alpha_n e^{\lambda_n t}$ (where the coefficients α_n are arbitrary complex, and the frequencies λ_n are arbitrary real) are quasi-periodic functions. According to a celebrated theorem by Harald Bohr, the class of almost periodic functions is identical to the closure $H\{s(t)\}$ of all finite $s(t)$. (The necessary and sufficient condition for an arbitrary trigonometric series, $\sum_{n=1}^{\infty} \alpha_n e^{i\lambda_n t}$, to be the Fourier transform of an almost periodic function is that $\sum |\alpha_n|^2$ converges, Besicovitch-Bohr).

We have that both “0” and “1” may be realized at each place S_i of the symbol string. Every refinement (sub-partitioning) of the binary symbolic dynamics will still lead to an expression for the topological entropy which is positive.

Somewhat analogously to the concept of a Lyapunov exponent (which is, roughly, an inverse correlation length in time) one would like to extract a “correlation length” ξ_{string} from the set of all allowed strings of symbols (30) measuring, in a quantitative way, that the correlation between the symbols (as extracted from the bounce structure captured from physical signals like $C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}(\tau)$) decreases with n as we go n steps along the “string” (from some i to $i + n$, averaged over i).

If all the realized string sequences (30) are *quasi-or-almost periodic*⁵⁰ (which we take to be a sign of integrability), we assign a correlation length of $\xi_{\text{string}} = \infty$. In a random sequence of symbols, on the other hand, there is no correlation between a symbol and the next, giving $\xi_{\text{string}} = 0$. If a correlation length can be defined for a string displaying “deterministic chaos” it should have a correlation length in between these two extremes.

For a given string of symbols the following limit may exist

$$C(n) = \lim_{N \rightarrow \infty} \left\{ \frac{1}{2N} \sum_{i=-N}^N S_i S_{i+n} - \left(\frac{1}{2N} \sum_{i=-N}^N S_i \right)^2 \right\}$$

and defines a correlation function $C(n)$ for the string S (i.e. along the string of symbols encoding a particular mixmaster gravitational collapse). If we can find just *one* symbol sequence for which the correlation function decays exponentially with n ,

$$C(n) \sim e^{-n/\xi}$$

it is not possible (we suspect) for the system to be integrable.

In fact, the mixmaster collapse orbit may very well (within the validity of the one-perturbation treatment of Belinskii et al.) be strongly intermittent in the “per bounce” time “ n ” and thus the correlation function will exhibit *power law decay* rather than decay exponentially along the string of symbols.⁵¹ Hence we may end up with a power law decaying correlation function, yet a positive topological entropy, for the infinite symbol strings of the mixmaster collapse orbits.

We shall not go into further details here. We remark that, starting out from our toy-model gravitational collapse (3), one is apparently naturally driven into speculations of a general sort in complexity theory. A general analysis of these problems is at the “heart of complexity theory”⁵² and is an extremely difficult subject.

3.8. Wheeler’s superspace metric as distance measure on the space of three-metrics?

If we wish to operate with a distance measure $\|\cdot\|$ to the space of three metrics, it is natural to investigate whether we have distance measures in general relativity which are more *natural* than others. In particular one would like to construct a distance measure which is invariant under (a large class of) changes of coordinates for the description.

⁵⁰“Quasiperiodicity” is captured by making a Fourier-transform on the symbol sequence.

⁵¹I thank Gábor Vattay, p.t. at the Niels Bohr Institute, for fruitful exchanges concerning this point

⁵²See also sec.5.4 in P. Grassberger *et al* (1991) for references to a number of different definitions of complexities of time sequences which are all tied to symbolic dynamics. Interesting discussions of complexity, in a wider perspective, may also be found in P. Grassberger (1986) and e.g. S. Lloyd and H. Pagels (1988).

The original dreams by Wheeler, DeWitt and others is to consider the dynamics of the three metrics $(^3)g$ as geodesics on some manifold called superspace equipped with the metric tensor G_{ijkl} which is named the supermetric. This supermetric G_{ijkl} induces a norm on the space of three metrics $(^3)g$,

$$||\delta g_{ab}||^2 = \int d^3x \sqrt{g} G^{ijkl} \delta g_{ij} \delta g_{kl} = \int d^3x \sqrt{g} G^{AB} \delta g_A \delta g_B \quad (31)$$

and one may measure the distance between two three-metrics $(^3)g$ and $(^3)\tilde{g}$ with respect to the G_{ijkl} tensor. More explicitly, one read off the first natural candidate for a metric on the configuration space (the space of three metrics) from the structure of the kinetic term in the ADM-Hamiltonian

$$H = \frac{1}{2} G_{(ij)(kl)} \pi^{ij} \pi^{kl} - \frac{1}{2} g^{(3)} R = \frac{1}{2} G_{AB} \pi^A \pi^B - \frac{1}{2} g^{(3)} R = 0 \quad (32)$$

which yields an expression like⁵³

$$G_{AB} \equiv G_{(ij)(kl)} = \frac{1}{2} (g_{ik} g_{jl} + g_{il} g_{jk} - 2g_{ij} g_{kl}) . \quad (33)$$

This step corresponds to reading off the “usual Euclidian metric” a^{ij} from a non-relativistic Hamiltonian⁵⁴

$$H = \frac{1}{2} a^{ij} p_i p_j + V(\mathbf{q}) = E . \quad (34)$$

Some may stop at this point and say that G_{ijkl} is the superspace metric. However the dynamics of the three-metric $(^3)g$ (the collapse orbit) is not yet a geodesic flow (cf. Biesiada and Rugh (1994)) with respect to this superspace metric.

By conformal transformation of the metric

$$\tilde{G}_{AB} = 2(E - V)G_{AB} = (-2V)G_{AB} = \mathcal{R}G_{AB} = g^{(3)}R G_{AB}$$

(where the role of the potential V in the general relativistic context is played by the quantity $V = -\frac{1}{2}g^{(3)}R \equiv -\frac{1}{2}\mathcal{R}$) and rescaling of the parametrization of the collapse orbit (cf. detailed discussion in Biesiada and Rugh (1994))

$$d\tilde{\lambda} = 2(E - V)d\lambda = (-2V)d\lambda = \mathcal{R}d\lambda = g^{(3)}R d\lambda \quad (35)$$

one obtains (cf. also e.g. C.W.Misner (1972)) that, with respect to this new parameter $\tilde{\lambda}$ and with respect to the conformally rescaled metric \tilde{G}_{AB} , the three-metric $(^3)g$ is now an affinely parametrized geodesic

$$\frac{d^2 g^A}{d\tilde{\lambda}^2} + \tilde{\Gamma}_{BC}^A \frac{dg^B}{d\tilde{\lambda}} \frac{dg^C}{d\tilde{\lambda}} = 0 . \quad (36)$$

⁵³Note that the supermetric, in this notation (cf. C.W. Misner (1972)) differs by a conformal factor (\sqrt{g}) from the expression by DeWitt (DeWitt (1967)). This is allowed since the ADM action is invariant with respect to conformal transformations. I do not want to get entangled in too much detail concerning this point but refer to M. Biesiada and S.E. Rugh (1994).

⁵⁴Instead of an analogy with a free particle in special relativity, with Hamiltonian $H = \frac{1}{2}(\eta^{\mu\nu} p_\mu p_\nu + m^2)$ (cf. C.W. Misner (1972), p. 451) we want to point to (cf. Biesiada and Rugh (1994)) the complete one-to-one correspondence between construction of the superspace and the dynamics of a non-relativistic particle with the Hamiltonian $H = \frac{1}{2}a^{ij} p_i p_j + V(\mathbf{q})$ reduced to geodesic flow by virtue of the Maupertuis principle.

There is no “magic” involved (we have “magic without magic”): Information about the potential $V = -\frac{1}{2}g^{(3)}R$ has simply been completely encoded in the mathematical definition of the superspace-metric \hat{G} which generates the evolution of ${}^{(3)}g$ as geodesic motion (affinely parametrized if one uses the rescaled parameter $\tilde{\lambda}$).⁵⁵

- The distance measures induced by the Wheeler superspace metric G_{ijkl} have the *good* property (which is not shared by artificially introduced Euclidean distance measures) of being invariant under canonical coordinate transformations. Thus, if we have a change of coordinates (configuration space variables),

$$g^A \rightarrow g^{*A} = g^{*A}(g^A)$$

the distance measure $ds^2 = G_{AB}\delta g^A\delta g^B$ is invariant under such transformations, since G_{AB} transforms properly as a tensor,

$$G_{AB}^* = \frac{\partial g^C}{\partial g_A^*} \frac{\partial g^D}{\partial g_B^*} G_{CD}. \quad (37)$$

One sees this by recalling that the coordinate transformation $g^A \rightarrow g^{*A}$ induces the transformation of momenta $\pi_A \rightarrow \pi_A^* = \frac{\partial g^B}{\partial g^{*A}} \pi_B$. Therefore the Hamiltonian \mathcal{H} reads

$$\mathcal{H} = \frac{1}{2}(G^{AB}\pi_A\pi_B - \mathcal{R}) = \frac{1}{2}(G^{AB}\frac{\partial g^{*C}}{\partial g^A} \frac{\partial g^{*D}}{\partial g^B} \pi_C^*\pi_D^* - \mathcal{R}) = \frac{1}{2}(G^{*AB}\pi_A^*\pi_B^* - \mathcal{R})$$

justifying our formula (37).

- A bad property of the distance measure induced by the Wheeler superspace metric is that such a distance measure is *indefinite* (not positive definite). I.e., it appears that one may have situations in which two spacetime metrics (two mixmaster collapses at some given “time”) are at *zero* distance but evolve into a distance different from zero. (This would correspond to a “Lyapunov exponent” of “ ∞ ” which is of course an entirely different situation from having a small finite distance which in “time” evolves into a bigger finite distance).

Note, that it *is possible* to construct a superspace metric G_{AB} which is *positive definite* so that two 3-geometries ${}^{(3)}g$ and ${}^{(3)}\tilde{g}$ are identical if and only if the distance between them is zero, see e.g. discussions in B.S. DeWitt (1970). However, such superspace metrics do not make the 3-geometries evolve as geodesics.

Negativity of the Ricci scalar $R < 0$ as a local instability criterium?

As a possible criterium for instability of the mixmaster gravitational collapse, M. Szydłowski and Lapeta (1990) and M. Szydłowski and M. Biesiada (1991) proposed to look at the Ricci scalar of the manifold on which the mixmaster dynamics generates a geodesic flow. (The Maupertuis principle was applied to the Hamiltonian formulation of the mixmaster dynamics as given in Bogoyavlensky (1985)) and investigate if one could extract (coordinate invariant) information about instability properties of the gravitational collapse in this way.

Application of the Maupertuis principle to the Bogojavlenskii Hamiltonian formulation of the mixmaster dynamics and the associated induction of a natural distance

⁵⁵In a completely similar manner (cf. Biesiada and Rugh (1994)) one maps the non-geodesic motion of the configuration space variable q governed by the non-relativistic Hamiltonian (34) to affinely parametrized geodesic motion by conformal transformation $a_{ij} \rightarrow 2(E - V(q))a_{ij}$ of the metric a_{ij} and rescaling of the time parameter $dt \rightarrow 2(E - V(q))dt$.

measure on the three metrics, is (cf. M. Biesiada and S.E. Rugh (1994)) exactly to implement the dream by Wheeler (described, e.g., in C.W.Misner (1972) for the mixmaster metric) to have a superspace metric “ G_{ijkl} ” (we call it Wheelers superspace metric) - with respect to which the three metrics move along geodesics - and use that metric as a natural distance measure on the space of mixmaster three-geometries. Thus, the Ricci scalar calculated from the Hamiltonian introduced by Bogoyavlenskii, cf. e.g. Szydłowski and Biesiada (1991), is exactly to calculate the Ricci scalar of the Wheelers superspace metric and the criterion $R < 0$ implies that one has *local* instability at least in one direction (as given by the *geodesic deviation equation* for two nearby mixmaster metrics investigated with Wheelers superspace metric as a distance measure).

An obstacle to this approach (of using the conformally rescaled Wheeler superspace metric), besides the obstacle that Wheelers superspace metric is indefinite, is the accompanying introduction of a host of singularities of the superspace metric which makes the original dream of Wheeler troublesome to achieve (i.e. the three metrics ⁽³⁾ g be geodesics w.r.t. a superspace metric) - even in the restricted class of mixmaster three geometries. Invariants calculated from the superspace metric, e.g. the Ricci scalar, inherit the singularities of the Wheeler superspace metric. In particular, the Ricci scalar quoted e.g. in M. Szydłowski and M. Biesiada (1991) has such singularities. This obstacle has also recently been emphasized by A. Burd and R. Tavakol (1993) and was previously discussed in C.W. Misner (1972).

However, as a *local* instability criterium, $R < 0$ indicates - at nonsingular points where it is defined - the *local* exponential instability in some directions of the configuration space w.r.t. the indefinite superspace metric.

Only insufficient attention, in my opinion, has been paid towards the often rather arbitrary distance measures on the space of three-metrics which are introduced in previous studies calculating Lyapunov exponents, etc.

It thus seems worthwhile to investigate if it is possible to construct and use good “gauge invariant” distance measures (for example the Wheeler superspace metric G_{ijkl} , with or without conformal transformation of G_{ijkl} , or other distance measures) on the solution space of three-metrics and if one can use such distance measures to discuss instability properties of e.g. the gravitational collapse in a way which is invariant under some (large) class of gauge transformations.

3.9. The mixmaster collapse as an interesting laboratory for testing ideas about how to apply chaos concepts in the context of general relativity

The structure of the mixmaster gravitational collapse model is so simple⁵⁶ that it is not unreasonable to investigate (cf. e.g. Contopoulos *et al* (1993)) whether the model should turn out to be integrable, after all. I.e., there could be several additional “hidden” symmetries in the governing Einstein equations of motion besides the Hamiltonian.

*A model like the mixmaster model collapse is very well understood in terms of the combinatorial model by Belinskii *et al* (controlled by simple maps). However, if we cannot agree on how to construct indicators of chaos (from*

⁵⁶In the approximation where the walls are infinitely hard we have an example of so-called “algebraic chaos”. All the interesting dynamics (the chaos) lies in the transition algebra associated with the bounces against the wall boundaries.

the full phase flow) in this simple example, how could we then, in principle, dream about constructing chaos indicators which can deal with more complicated situations?

Whether we can assign any invariant meaning to a concept like “metrical chaos” (or even more ambitious: “turbulence in spacetime metrics”) is a theoretical problem of intrinsic interest in the mathematical study of the dynamics of the full non-linear Einstein equations.

4. IS “CHAOS” A GENERIC FEATURE OF THE EINSTEIN EQUATIONS?

The case study of our mixmaster metric (3) shows that the “generic” expectation of the self-interacting gravitational field to generate space-time “chaos and turbulence” (for strong field strength, e.g. when probing the Einstein equations near space-time singularities where invariants like $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$ of the tidal field curvature tensor blow up without upper bound) is reflected even in such a *highly symmetric* model. This may be considered a very strong property:

If such a “simple” (finite degrees of freedom) deterministic system shows a chaotic, complex dynamical behavior, then one expects this to be even more the case in complicated scenarios with a larger number of interacting degrees of freedom!

In this sense the “mixmaster” toy model of a gravitational collapse should be considered as a promising starting point of subsequent investigations of more complicated situations (with less symmetry) of the classical field equations in general relativity.

Compare, e.g. with the famous Lorentz attractor model (Lorentz, 1963) for turbulence which is a crude starting point for an investigation of some properties of hydrodynamical turbulence. Note, however, that the Lorentz attractor is a highly truncated model, where only very few modes of the Fourier expanded equations are kept in the model; The governing equations for the mixmaster, $SU(2)$ homogeneous, metric are *the exact Einstein equations* - involving few degrees of freedom because of the (a priori) made symmetry-ansatz of the metric.

4.1. Aspects of fragility of the gravitational collapse?

However, also for governing equations of motion which are modified slightly relative to the Einstein equations one would expect that “metrical chaos” is a generic feature when probing the metric in regions where the curvature gets strong and non-linear self-interactions of the involved gravitational fields get important (e.g. on approach to curvature singularities).

Let us say that we modify the Einstein-Hilbert action slightly,

$$S = \int d^4x \sqrt{-g} \{ R + \epsilon \Phi(R, R_{\mu\nu}R^{\mu\nu}, R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}) \} \quad (38)$$

where Φ is a general analytic function and ϵ is a small perturbation parameter. See also discussions in R.K. Tavakol (1991) and A.A. Coley and R.K. Tavakol (1992).

Chaos in the evolution of the metric (3) should not merely be a property of the *exact* Einstein equations (i.e. for $\epsilon = 0$), but also for the modified equations, corresponding to the inclusion of higher order curvature terms (i.e. for $\epsilon \neq 0$) in the action. Such higher order curvature terms - generically arising in underlying theories of gravity like, for example, string theories - will *blow up* towards the ultraviolet (for example towards the big crunch singularity of the mixmaster metric).

Thus, when curvature gets big there is no reason why higher order curvature terms should not appear in the Lagrangian, $\mathcal{L} = \sqrt{-g} \{R + \alpha_1 R^2 + \dots\}$. Seen in this perspective, it is puzzlesome that the chaotic mixmaster oscillations are unstable and stop at some stage of the collapse (and proceed in a non-oscillatory manner) if higher order curvature terms, like R^2 terms, are included in the Lagrangian, according to J. Barrow and M. Sorousse-Zia (1989). See, also, J. Barrow and S. Cotsakis (1989) and a recent study by S. Cotsakis *et al* (1993).

A (premature) conclusion arises, that the mixmaster oscillatory behavior - or the very fact that the metric (3) is evolving in a complex (unpredictable) manner towards the big crunch singularity - is *not* structurally robust towards modifications of the Einstein-Hilbert action and such modifications are certainly expected when length scales get small, of the order of Planck sizes, say.

4.2. A remark on fragility and structural stability in a wider context

In a broader context R. Tavakol (1991) discusses the concept of “structural *fragility*” and Coley and Tavakol (1992) advocate that models which are structurally fragile are “generic” among the set of models usually employed in cosmology.

We have already mentioned that isotropization of anisotropic spacetime metrics like the compact mixmaster metric (3) occurs too slowly within the standard hot big bang model to explain the remarkable degree of isotropy in the microwave background radiation observed today. Thus, isotropy is an unstable property of physically realistic cosmological initial conditions - yet it is an observed property of the Universe. Another, famous, example is the “flatness” problem (R.H. Dicke): Why is the Universe today so near the boundary between open and closed, i.e. so nearly flat? This is a puzzle because one may show that $\Omega = \rho/\rho_c = 1$ is an *unstable equilibrium point* of the evolution of the standard hot big bang theory (i.e. it resembles the situation of a needle balancing vertically on its tip). In order for Ω to be somewhere in the allowed range today, $\Omega \sim 1/10 - 10$, this parameter had to be equal to one to an accuracy of 49 decimal places if we consider times around 10^{-35} second, say, after the big bang! This finetuning problem is, however, no problem at all relative to the completely vanishing probability for us to have a second law of thermodynamics which R. Penrose (1989) estimates as one chance in $\sim e^{10^{123}}$! (The Universe was created in a state of very low entropy compared to what it might have been).

These finetuning problems are to some extent relaxed within the inflationary Universe concept. Inflation predicts, for example, that the value of Ω today should equal one to an accuracy of about one part in 100.000, cf. A.H. Guth (1992).

In view of the variety of simplifying assumptions in the formulation of any mathematical model of a physical phenomenon in Nature and the fact that such models have a tremendous power in describing such phenomena, it appears that such a principle of “structural fragility” cannot be implemented in Nature to its extreme limit.

The extreme opposite principle, the so-called “random dynamics” principle, which concerns the structure of Natural laws themselves, has been put forward by Holger Bech Nielsen (at the Niels Bohr Institute). He contemplates (cf. e.g. H.B. Nielsen (1976, 1981)) - and has explored in many contexts (cf., e.g., C.D. Froggatt and H.B. Nielsen (1991)) - a principle of the (extreme) structural stability of the Natural laws (as they appear to us) - against (almost any) modification of them in the ultraviolet.

It is a principle which postulates the structure of Natural laws (at our scales) despite a substantial lack of structure of laws at a more fundamental level and the dream, though very ambitious here, has some analogy to, say, the phenomenon of “universality in chaos”, i.e. the realization that many phenomena, like the universality of Feigenbaum’s δ in a bifurcation scenario, do not hang severely on the *microscopic details* of the underlying dynamics.⁵⁷ In the context of our discussion of the mixmaster collapse it also reminds us of the “chaotic cosmology” concept (see also sec.6.3), developed by C.W. Misner, which attempts at creating our present Universe from (almost) arbitrary initial conditions.

If the universality class of fundamental models which leads to the same infrared phenomenology (the standard model of the electroweak and strong forces, say) is not too narrow it leaves (logically) the possibility of having “*chaos in the Natural laws*” (at the most fundamental “scales”) - i.e. the fundamental structure may be selected randomly if it is just selected from the universality class of models restricted by the boundaries set by the “universality class”

However, such boundaries (if these could be made precise enough to be stated formally, say, by mathematical formulas) in a certain sense themselves have status as “Natural laws” (regularities) - and so on, ad infinitum. Therefore, this way, one does not circumvent the concept of some “Natural laws” and “regularities” to be implemented in our Universe.

Preliminary discussions of boundaries that such a “random dynamics” principle may come across (is it possible to define a concept of “structural stability for Natural laws”?) is contained as sec.3 in H.B. Nielsen and S.E. Rugh (1993). Boundaries that such a project come across are of interest because hereby one may gradually try to isolate elements in our description of Nature which can not (so easily) be modified (and thought differently). If one contemplates - cf. e.g. S. Weinberg (1992) - that the Natural laws and parameters are the only *self-consistent* set of laws and parameters imaginable, the “random dynamics” project contributes with an analysis of what we could mean by the word “self-consistent”.

4.3. Back to spacetime chaos: “Chaos” in other solutions in classical general relativity?

It would be a good idea to look for other toy models of spacetime metrics, which would exhibit chaotic behavior, e.g. in connection with solutions obtained in the (fast advancing) discipline of *numerical relativity* - if one has good “codes” available to solve the Einstein equations without too many symmetry assumptions!

If possible, it would be very nice if one could relax the symmetry properties and study the structure of the cosmological singularities of space-time metrics with only two spacelike Killing vector fields, say, instead of three. Especially, it would be wonderful

⁵⁷I.e. there is a huge universality class of dynamical systems, e.g. all one-dimensional maps with a quadratic maximum (cf. also Cvitanovic (1989)), which generates the same Feigenbaum δ .

if one could relax the symmetry properties and generalize the mixmaster gravitational collapse this way.

It is a reasonable expectation that the severe non-linearity present in the Einstein equations implies “chaotic” (non-integrable) solutions almost “unavoidably” in scenarios involving high gravitational field strength (e.g. near space-time singularities). This “metrical chaos” should therefore be thought of as a generic feature of the non-linear Einstein equations in such strong gravitational fields - and not as a feature connected, merely, to the near-singularity behavior of some (small) subclass of metrics, i.e. the spatially homogeneous toy-models of gravitational collapses.

On the “universality class” for the intermittent behavior of the “u” parameter (discussions with V.N. Lukash)

Even in the (small) class of spatially homogeneous metrics we have, according to Peresetzki and V.N. Lukash, cf. Lukash (1983), as regards the evolution of the already mentioned “Lifshitz-Khalatnikov” parameter “u”, that their evolution is chaotic (and actually given by the Farey/Gauss map) if the invariance groups/algebra’s - according to the standard classification of three-dimensional Lie-algebras by Bianchi (1897) - (of the three-spaces) are chosen as type *VIII* and *IX* (type *IX* is actually the mixmaster space) in the case of the *empty space* equations $R_{\mu\nu} = 0$, but in fact *all the Bianchi metrics* - except for type *I* (a flat space) and *V* (where the curvature is isotropic) - in situations (a perfect fluid energy momentum tensor is assumed) where $T_{\nu\nu} \neq 0$ (but $T_{ij} = 0$), cf. Peresetzki and V.N. Lukash.

In fact, one can show that the spatially homogeneous toy-model metrics are driven into “chaos”, as regards the evolution of the parameter “u”, by the presence of *curvature anisotropy of spiral character* (spiral character: At least one structure constant is different from zero, $C_{jk}^i \neq 0$, where all i, j, k are different, in the Lie-algebra formed by the three spacelike Killing vector fields); For a space of type *I* there is no curvature at all, for a metric of type *V* it is isotropic and - as a curiosum - for a space of Kantowski Sachs type, the anisotropic curvature is not of spiral character! (V.N. Lukash, private comm., NBI, and V.N. Lukash, Doctorate Thesis (1983)).

• Solutions with other “matter” sources:

What if other sources than perfect fluid matter are included in the study? It is known, e.g., that the $SU(2)$ Yang Mills equations in the limit of infinite wavelength (i.e. the spatially homogeneous mode) admit “chaotic” solutions⁵⁸ (cf., e.g., S.G. Matinyan (1985) and G.K. Savidy (1984)) with a positive metric entropy $h = \int \sum \lambda^+ d\mu > 0$. (In this limit - the “extreme infrared” - it is also a famous result of linear stability analysis that the non-Abelian Yang-Mills configurations are *unstable* under small disturbances, cf. N.K. Nielsen and P. Olesen (1978) and S.J. Chang and N. Weiss (1979).

If the Einstein and $SU(2)$ Yang Mills equations are coupled, e.g. in the context of a mixmaster toy model study of a gravitational collapse, some very complicated and chaotic behavior must arise!?⁵⁹

⁵⁸For more recent discussions of spatially inhomogeneous chaos in non-Abelian gauge theories see also e.g. M. Wellner (1992), B. Müller *et al* (1992) and T.S. Biró *et al*. (1993). A detailed description of chaos in non-Abelian gauge theories will also be found in S.E. Rugh (1994).

⁵⁹For a study of stability properties of such coupled $SU(2)$ Einstein-Yang-Mills-Higgs equations (in the context of monopole and black hole solutions) see also Straumann and Zhou (1990) and G.W. Gibbons (1990).

- **Spatially inhomogeneous gravitational collapses:**

What is the effect of taking into account spatial degrees of freedom in the model; i.e., the inclusion of some higher order Fourier modes (“spin 2” gravitational waves) superimposed (in a gauge-invariant way?) on the $SU(2)$ homogeneous (zero-mode expanded) but anisotropic gravitational collapse? Perhaps small amplitude gravitational waves die out towards the singularity? It is, however, far from obvious that such perturbations will not do the opposite: Blow up and significantly alter the evolution of the metric on approach to the space-time singularity! (Cf, also, remarks by R. Ove (1990) in a slightly different context).

Note, that according to Belinskii, Khalatnikov and Lifshitz (1982), the simple BKL-combinatorial model for the alternation of Kasner exponents (derived for the spatially homogeneous models) remains valid locally (i.e. in the neighborhood of every space point) in the general spatially inhomogeneous gravitational collapse. See also discussion in Ya.B. Zel'dovich and I.D. Novikov (1983) §23.3.

- **Chaos in colliding gravitational waves?**

One may speculate whether yet another example of chaos in the classical Einstein equations occurs in the dynamics of the metric field (and extracted gauge invariant quantities) when two non-linearly interacting plane gravitational waves extending to infinity collide and generate black hole singularities - via *nonlinear focusing* effects - of the “spacetime” structure⁶⁰. The detailed nature of the evolutionary tracks of the gravitational field $\{g_{\mu\nu}(x)\}$ “ - when “bouncing off” parts of the spacetime manifold in the form of black hole singularities - must almost unavoidably (?) be *chaotically unpredictable!* Highly nonlinear effects (when the curvature gets strong) determines the detailed nature of the dynamics.

5. IMPLICATIONS OF “METRICAL CHAOS” ON THE QUANTUM LEVEL?

Since the classical Einstein equations are scale-independent, there is - a priori - no scale built in our toy-model study (3) of the oscillatory, chaotic collapse - when evolved according to the classical Einstein equations. However, if we - by hand - put in a length-scale (lets say “the size” of the Universe $\sim c/H \sim 10^{28} \text{ cm}$, or somewhat less) as a starting condition on the scale factors of the gravitational collapse, one notes (cf. also the displayed fig.5) that already after few major cycles of bounces some of the length scales of the gravitational collapse get very small and, in fact, very fast reach Planck-scales

$$l_{Pl} = \sqrt{\frac{\hbar G}{c^3}} \sim 10^{-33} \text{ cm}$$

and, correspondingly, the space-time curvature blows up

$$\sqrt{R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}} \sim l_{Pl}^{-2} = \frac{c^3}{\hbar G} \sim 10^{66} \text{ cm}^{-2} .$$

Such scales may set characteristic scales when quantum effects become important and the gravitational degrees of freedom should be quantized.

⁶⁰Cf., e.g., S. Chandrasekhar, talk at the “300 years of Gravity” celebrations at D.A.M.T.P. (1987). However, it should be realized that the self-focusing singularities most likely arise because they are colliding plane waves which extend to *infinity* - and such waves do not exist! (M. Demianski and D. Christodoulou, pers. comm.).

Thus, the mixmaster gravitational collapse turns very fast into a quantum problem! (and is chaotic in a domain where it should be quantized)

The fact that the scale functions for the mixmaster spacetime metric change so fast can be understood from the observation that the Einstein equations for the metric (3) involve exponential functions (cf. (7) or (25)) and such exponential functions are well known to give out “large numbers” very fast!

We are lead to consider our toy-model study of a chaotic gravitational collapse as:

1. Merely a look into a “chaotic sector” of the solution space to the *classical* Einstein equations. One notes that the classical Einstein equations do not internally lead to contradictions at any scale (see also Landau and Lifshitz, Vol.II, §119) and considered as a *mathematical study* of the non-linear Einstein equations, “metrical chaos” is a concept (not yet properly defined) of interest in itself!
2. Presenting some evidence in favor of speculations that the process of quantization of the gravitational field - occurring at small lengthscales and correspondingly high curvatures ($\sqrt{R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}} > l_{Pl}^{-2}$) - in generic situations rather will occur “around” field configurations which are “turbulent” (space-time chaos) than around well behaved (integrable) solutions⁶¹. In the semiclassical limit, which is built around the classical configurations, this may be of importance.

Does classical chaos obstruct the quantization of the theory? Lee Smolin (1990) makes the point that one will - very likely - only make progress on the construction of quantum general relativity to the extent that one uses *non-trivial information* about the *dynamics* of general relativity.

In perturbation theory and Monte-Carlo simulations (methods of quantization in which no special information about the dynamics of the theory is used) chaos usually does not obstruct the quantization. Ashtekhar and Pullin (1989), for example, emphasize that chaotic behavior of Yang-Mills fields has not obstructed the quantization of the theory. Such questions, being far from settled, are central topics in the theory of “Quantum Chaos” and will also be discussed (and references may be found) in S.E. Rugh (1994).

Spatially homogeneous cosmologies - having a finite number of (anisotropy) degrees of freedom - are often considered a laboratory for testing ideas about the application of quantum principles to gravitational degrees of freedom. Examples could be the Robertson Walker, the Kasner or the mixmaster spaces. Now, the Kasner or FRW solutions (being integrable) do not display any “chaos” at the classical level, while the evolution of the mixmaster metric is chaotically complicated (and in metric t time the dynamical evolution condenses, as regards the oscillations of the anisotropy degrees of freedom, infinitely towards the two spacetime singularities).

Does this striking difference in the dynamical behavior at the classical level mirror itself in quantum effects in the “quantized” model cosmologies? I.e. what is the effect of applying quantum principles to a chaotic, dynamically

⁶¹Thus, our “intuition” (on gravitational collapses) based on the well known textbook example of the FRW-type “big bang”/“big crunch” collapse - with simple dynamical behavior - is a non-typical classical field configuration to have in mind when “quantizing gravity”.

complicated gravitational collapse - the mixmaster, say - compared to more symmetric gravitational collapses (of the FRW- or Kasner-types, say) which are integrable all the way to the “final crunch” singularity?

In a *semiclassical* treatment (i.e. when the collapse has not yet reached Planck lengths) the quantum solution builds around the classical dynamics. So in the semi-classical limit it of course, for that reason, matters what the classical solution looks like!

What about the *fully quantized* regimes, when the length scales of the gravitational metric have collapsed down to Planck sizes, or less? It is hard to know what happens in the fully quantized regime of our gravitational collapse (despite that it reaches these small scales very fast) let alone that no good theory of quantum gravity is available.

5.1. The Wheeler-DeWitt quantization of the gravitational collapse

Since the pioneering works of P.A.M. Dirac in the fifties (cf., e.g., P.A.M. Dirac (1959) and references therein) and R. Arnowitt, S. Deser and C.W. Misner (1962), it is well known that a Hamiltonian formalism (ADM 3+1) may be set up for the Einstein equations and this Hamiltonian formalism is often used in a “toy-application” of quantum principles (the Wheeler-DeWitt equation) to such a model of an anisotropic gravitational collapse. See, e.g., the recent bibliographies on these topics by C. Teitelboim and J.J. Halliwell in S. Coleman *et al* (1992).

More explicitly, the Wheeler-DeWitt equation (J.A. Wheeler (1968), B.S. DeWitt (1967)) arises after appropriate translation of the coordinate and momentum variables into operators by imposing the Hamiltonian constraint⁶² as an operator constraint on the “quantum wave function Ψ ” of the gravitational collapse.

For the sake of notational simplification, we put $16\pi G = 1$, $\hbar = 1$, etc., and the Wheeler-DeWitt equation (the operator form of the Hamiltonian constraint) has the form

$$\hat{\mathcal{H}} \Psi = (G_{ijkl} \frac{\delta^2}{\delta g_{ij} \delta g_{kl}} + g^{1/2} {}^{(3)}R - {}^{(3)}\mathcal{L}_m) \Psi = 0 \quad (39)$$

where $\Psi = \Psi({}^{(3)}g) = \Psi(g_{ij})$ is a quantum mechanical wave-function (for our gravitational collapse), ${}^{(3)}R$ is the curvature scalar of the three-metric in question, ${}^{(3)}\mathcal{L}_m$ denotes the three-dimensional Lagrangian density of the (non-gravitational) matter fields and

$$G_{ijkl} \equiv \frac{1}{2} g^{-1/2} (g_{ik}g_{jl} + g_{il}g_{jk} - g_{ij}g_{kl})$$

is the Wheeler-DeWitt superspace metric (cf., e.g, discussion in B.S. DeWitt (1967)) on the space of the three-geometries.

A substantial amount of uncertainty could be expressed as to whether a toy-application of quantum principles along such lines makes good sense when implemented in a context like the mixmaster gravitational collapse (3). Some points of uncertainty:

⁶²Such constraints always arise in field theories where the field variables have a “gauge arbitrariness” (e.g. also in the Hamiltonian formulation of Yang-Mills theory) and the vanishing of a Hamiltonian, like in equation (39), is a characteristic feature of theories which are invariant under reparametrizations of time (cf., e.g. R.M. Wald (1984), appendix E). There is, however, a crucial difference between gauge theories and *parametrized* theories like general relativity (also invariant under reparametrizations which involve time). In a parametrized theory the *constraints* give all the *dynamics*. See also e.g. J.B. Hartle and K.V. Kuchař (1984).

1. The “factor ordering” ambiguity in translating a classical equation like the classical Hamiltonian constraint into an operator identity (involving non-commutating quantum operators) is a severe problem. For example, one may select a factor ordering - in the case of our mixmaster toy-model gravitational collapse - that makes *any given function* Ψ (of the mixmaster three-metric variables Ω, β_{\pm}) a solution of the Wheeler-DeWitt equation for this factor ordering! (cf. investigations by Kuchar quoted in Moncrief and Ryan (1991), p.2377).

2. It is not clear that quantum solutions of the highest-symmetry models are approximations to quantum solutions of models with less symmetry (e.g. that “quantizing” the spatially homogeneous mixmaster collapse is an approximation to some “true quantum gravity solution”, which includes space-dependent modes). Cf. Kuchar and Ryan (1989).

3. If space-dependent gravitational waves are superimposed on the mixmaster metric and taken into account perturbatively in the Wheeler-DeWitt equation one will most likely end up with non-renormalizable divergencies?

4. The symbol “ Ψ ” (the “wave-function” of the quantized gravitational collapse including matter degrees of freedom etc.) appearing in equation (39) does not have a straightforward interpretation if it also includes the observer, for instance in an application of quantum principles to an entire model-cosmology. One of the difficulties one encounters is how to allow for an experimentalist with a “free will” to perform experiments, if *everything* (including the experimentalist) is described by a big (completely deterministic) “Schrödinger equation”? The very notion of an experiment seems meaningless, since the measurement was determined in advance (at the “big bang”)?⁶³

In lack of a truly renormalizable underlying theory of quantum gravity, one may choose to ignore such obstacles and consider a toy-application of the Wheeler-DeWitt equation (39) implemented in the case of the mixmaster gravitational collapse and hope that it retains aspects of the full theory.⁶⁴

What do we obtain? We yield an operator constraint equation which roughly has a mathematical form as given below if we consider the simplest possible choice of factor ordering of the operators (we neglect non-gravitational matter fields)

$$\left\{ -\frac{\partial^2}{\partial\Omega^2} + \frac{\partial^2}{\partial\beta_+^2} + \frac{\partial^2}{\partial\beta_-^2} + e^{-6\Omega} {}^{(3)}R(\Omega, \beta_{\pm}) \right\} \Psi(\Omega, \beta_{\pm}) = 0 . \quad (40)$$

It looks like a *zero energy* “Klein-Gordon-Schrödinger” wave equation. Ω, β_+, β_- are the degrees of freedom in the parameterization (17) of the metric (3) and Ω, β_+, β_- are one-to-one related to the scale functions a, b, c in (3) via

$$\begin{aligned} a &= \exp(-\Omega + \beta_+ + \sqrt{3}\beta_-) \\ b &= \exp(-\Omega + \beta_+ - \sqrt{3}\beta_-) \\ c &= \exp(-\Omega - 2\beta_+) \end{aligned}$$

$\Omega = -\frac{1}{3} \ln(abc) \propto -\log(Volume)$ may function as a “time” parameter of the gravitational collapse since, as we have already noted, the three-volume $V = 16\pi^2 abc$ of the metric (3) is monotonically decreasing on approach to the space-time singularity and the variables $\beta_+ = \frac{1}{6} \ln(ab/c^2)$, $\beta_- = \frac{1}{2\sqrt{3}} \ln(a/b)$ denote the “state of anisotropy” of

⁶³See also, e.g., chapt. 15 “Quantum Mechanics for Cosmologists” in J.S. Bell (1987).

⁶⁴According to R. Graham (1993): “Given a dynamical system of Hamiltonian form, the temptation to quantize seems to be irresistible”.

the gravitational collapse. The three-curvature scalar on the three-space, ${}^{(3)}R$, can be calculated from the metric (3) and has the following form

$$\begin{aligned} {}^{(3)}R &= -e^{2\Omega} \left\{ e^{4\beta_+} (\cosh(4\sqrt{3}\beta_-) - 1) + \frac{1}{2} e^{-8\beta_+} - 2e^{-2\beta_-} (\cosh(2\sqrt{3}\beta_-)) \right\} \\ &= -\frac{1}{2a^2 b^2 c^2} \{(a^2 - (b+c)^2)(a^2 - (b-c)^2)\} . \end{aligned} \quad (41)$$

The structure of ${}^{(3)}R$ gives rise to a scattering potential (in the initial stages of the collapse describing soft scattering walls which fast develop, however, into “infinitely hard” walls when approaching the space-time singularity of the metric) which makes the classical Hamiltonian dynamical problem resemble that of a billiard ball being played on a billiard table (the anisotropy plane $(\beta_+, \beta_-) \in \mathbf{R}^2$) with this potential as the boundary walls. Of course, in the isotropic case $a = b = c = R/2$ we reobtain the expression for the three-curvature scalar ${}^{(3)}R = +3/2a^2 = +6/R^2$ corresponding to the compact ($k = +1$) *FRW* space.

The Wheeler-DeWitt quantized mixmaster collapse has been studied by various authors, for example C.W. Misner (1972) (see also, e.g., S.W. Hawking and J.C. Luttrell (1984), R. Graham and P. Szépfalusy (1990) and V. Moncrief and M.P. Ryan (1991), and references therein, for a toy “quantization” of the mixmaster metric).

Recently, exact solutions have been found by R. Graham if an additional supersymmetry is introduced in the study, R. Graham (1991, 1992). (See, also, P.D. Eath, S.W. Hawking and O. Obregón (1993)). These solutions describe virtual quantum wormholes, see also R. Graham (1993).

In the context of the Wheeler-DeWitt equation, applied to spatially homogeneous gravitational collapses, we could hope (preliminarily) to address the question: What is the effect of having a chaotic gravitational collapse (as opposed to an integrable solution) when quantizing it? Since in the Wheeler-DeWitt equation we are merely looking at the zero energy solution $\hat{\mathcal{H}}\Psi = 0$, it is at first sight not obvious how we can relate to, say, an effect (cf., e.g., M.V. Berry (1981)) like that of nearest neighbour energy repulsions. Prof. R. Graham, however, pointed out that C.W. Misner already twenty years ago offered a separation of the Wheeler-DeWitt equation in the region sufficiently near the space-time singularity where the scattering potential $g\,{}^{(3)}R$ is approximately infinitely hard and where the solutions to the Wheeler-DeWitt equation are related to the spectral properties of the Laplace operator on the Poincaré disc.

Before I sketch this last point I will digress shortly into a description of the mixmaster gravitational collapse of remarkable beauty.

5.2. The Poincaré disc description of the gravitational collapse

The dynamics may be transformed ⁶⁵ into that of piecewise geodesic motion in a non-compact domain with triangular symmetry (corresponding to the symmetry under the interchange $a \leftrightarrow b \leftrightarrow c$ of the scale factors in the metric (3)) on the two dimensional Poincaré disc (the Lobachevsky space) of constant negative curvature! As is well known, geodesic motion on surfaces of constantly negative curvature is a standard laboratory for testing ideas in classical and quantum chaos, cf. e.g. Balazs and Voros (1986). However, the mixmaster gravitational collapse may very well be one of the only physically

⁶⁵Note, that the Poincaré disc description of the mixmaster collapse has gone almost unnoticed for twenty years. See, however, J. Barrow (1982). Suddenly several people looked at it again, J. Pullin (1990), Graham and Szépfalusy (1990), (S.E. Rugh (1991)).

motivated⁶⁶ dynamical systems known, which realizes such geodesic motion on surfaces of constantly negative curvature! It is thus clearly an interesting and beautiful aspect of the mixmaster collapse which deserves further investigations. (The Poincaré disc description has been emphasized also by J. Pullin, recently, cf. J. Pullin (1990) and R. Graham and P. Szépfalusy (1990)).

A very brief discussion of the set of transformations which brings the mixmaster gravitational collapse into that of geodesic motion on the Poincaré disc is offered in Misner, Thorne and Wheeler (1973) §30.7. I will not drift into details here - since that is far beyond the scope of this presentation (a detailed description will be available in S.E. Rugh (1994), where the “Poincaré disc” description is worked out and discussed in exhaustive detail).

The crucial observation (which is originally due to C.W. Misner and D.M. Chitre) is that while the original Hamiltonian

$$\mathcal{H} = -p_\Omega^2 + p_+^2 + p_-^2 + g^{(3)}R(\Omega, \beta_+, \beta_-) \quad (42)$$

has outward, time dependent expanding potential boundaries in the Ω, β_{\pm} variables, a set of transformations

$$(\Omega, \beta_+, \beta_-) \leftrightarrow (t, \xi, \phi) \leftrightarrow (t, x, y)$$

may be devised

$$\begin{pmatrix} \Omega - \Omega_0 \\ \beta_+ \\ \beta_- \end{pmatrix} = e^t \begin{pmatrix} \cosh \xi \\ \sinh \xi \cos \phi \\ \sinh \xi \sin \phi \end{pmatrix} = e^t \frac{1}{1 - (x^2 + y^2)} \begin{pmatrix} 1 + (x^2 + y^2) \\ 2x \\ 2y \end{pmatrix} \quad (43)$$

which makes the location of the potential boundaries *time-independent* with respect to the new “ t ” time parameter when we are in a region sufficiently near the space-time singularity of the metric. (The second part of the transformation (43) is a mapping⁶⁷ of (ξ, ϕ) into coordinates (x, y) inside the Poincaré unit disc \mathcal{D} in the complex plane, see below).

The transformations of the momenta p_Ω, p_+, p_- which appear in the Hamiltonian (42) are constructed to make the transformations of coordinates and momenta become *canonical* transformations, $p_i' = (\partial q^i / \partial q^i') p_i$, and the Hamiltonian is (after proper rescalings) transformed into a Hamiltonian which, in (t, ξ, ϕ) coordinates, has the form

$$\tilde{\mathcal{H}} = -p_t^2 + p_\xi^2 + \frac{p_\phi^2}{\sinh^2 \xi} + \tilde{V}(t, \xi, \phi). \quad (44)$$

The scattering potential $\tilde{V}(t, \xi, \phi)$ has (in the asymptotic region sufficiently near the singularity of the metric) infinitely steep potential boundaries and the location of these potential boundaries (which show up to form a triangular domain $\mathcal{B} \subset \mathcal{D}$ inside the Poincaré disc, see below) are *independent* of the new “time” variable t .⁶⁸ The location

⁶⁶If this system is not a “physical system” in an *empirical* sense, it is certainly a physically motivated system, being a (globally) perturbed FRW metric, evolved by the full, non-linear Einstein equations.

⁶⁷There are some discrepancies with the expressions in J. Pullin (dec. 1990). I, however, get these set of transformations.

⁶⁸Note, according to (43), that the new “time” parameter t is a somewhat “strange” new time variable,

$$t = \frac{1}{2} \ln((\Omega - \Omega_0)^2 - (\beta_+^2 + \beta_-^2)) = \frac{1}{2} \ln((\Omega - \Omega_0)^2 - ||\vec{\beta}||^2)$$

which mixes the state of anisotropy $\vec{\beta}$ (the position of the “billiard ball”) with the original Ω -time variable, $\Omega = -\frac{1}{3} \ln(abc)$, in the ADM-Hamiltonian variables.

of the potential boundaries are in the (ξ, ϕ) space determined by the equations, cf. MTW (1973) §30.7,

$$2 \tanh \xi = -\frac{1}{\cos(\phi + m \frac{2\pi}{3})} , \quad m = -1, 0, 1 . \quad (45)$$

Having obtained in this way the stationary (in the (ξ, ϕ) plane) scattering potential $\tilde{V}(\xi, \phi)$ which is zero inside the domain (we call this domain \mathcal{B}) bounded by (45) and ∞ in the region outside this domain (an approximation which is extraordinarily good when sufficiently near the “big crunch” singularity of the mixmaster gravitational collapse)

$$\tilde{V}(t, \xi, \phi) \rightarrow \tilde{V}(\xi, \phi) = \begin{cases} 0 & \text{inside domain } \mathcal{B} \\ +\infty & \text{outside domain } \mathcal{B} \end{cases} \quad (46)$$

we have obtained a Hamiltonian (44) which, in the interior of the scattering domain \mathcal{B} , resembles the Hamiltonian for a free particle⁶⁹ which moves along the geodesics on a curved manifold with constant negative curvature.

To see this, note, that in general the Hamiltonian for a particle that moves along the geodesics ($\delta \int ds = 0$) of a manifold with the metric $g_{\mu\nu}$,

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \quad (47)$$

has the form

$$\mathcal{H} = \frac{1}{2m} g^{\mu\nu} p_\mu p_\nu \quad (48)$$

where the metric tensor $g^{\mu\nu}$ is the inverse of $g_{\mu\nu}$, i.e. $g_{\mu\nu} g^{\nu\rho} \equiv \sum_\nu g_{\mu\nu} g^{\nu\rho} = \delta_\mu^\rho$.

Inside the (stationary) potential walls (46) we have arrived at the stationary Hamiltonian (44). As regards the *projected* motion on the two dimensional plane (ξ, ϕ) , we may therefore conclude from the form of the Hamiltonian (44) that the associated flow is a geodesic flow on a Riemannian manifold with the metric

$$ds^2 = d\xi^2 + \sinh^2 \xi d\phi^2 \quad (49)$$

inside the (stationary) region (the “billiard table” domain \mathcal{B}) bounded by infinitely steep potential walls (45). This metric (49) is the metric of the Poincaré disc of constant negative curvature! (cf. also, e.g., Balazs and Voros (1986)). Explicit computation of the Gaussian curvature K at the point (ξ, ϕ) gives

$$K = -\frac{1}{\sinh \xi} \frac{\partial^2}{\partial \xi^2} \sinh \xi = -1$$

As is well known, by suitable transformations of the coordinates, the geodesic flow on the Riemannian manifold (49) may be represented in various (equivalent) ways. Among these equivalent models (representations) we will consider the “Poincaré unit-disc model” for the hyperbolic geometry. By identifying a point (ξ, ϕ) with the point $z = (x, y) = r(\cos \phi, \sin \phi) \equiv re^{i\phi} \in \mathbf{C}$ (in the complex plane) via the coordinate transformation

$$z = x + iy = re^{i\phi} = \tanh(\xi/2) e^{i\phi} , \text{ i.e.} \quad (50)$$

⁶⁹Since the Hamiltonian (44) ceases to be time-dependent in the asymptotic region near the spacetime singularity, we have $\dot{p}_t = -\partial \tilde{H}/\partial t = 0$. Hence, p_t is a constant of motion and the bouncing of the (ξ, ϕ) variables (or the (x, y) variables (51)) within the scattering domain \mathcal{B} will take place at *constant speed* with respect to the t time coordinate.

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix} = \begin{pmatrix} \tanh \xi/2 \cos \phi \\ \tanh \xi/2 \sin \phi \end{pmatrix} \quad (51)$$

we obtain a hyperbolic flow (of constant negative curvature $K = -1$) on the unit disc

$$\mathcal{D} = \left\{ z = x + iy \in \mathbf{C} \mid |z| = \sqrt{x^2 + y^2} \leq 1 \right\}. \quad (52)$$

Any point of the (ξ, ϕ) space is transformed into the unit disc (52). The boundary (i.e. points at $r = 1$) of the disc \mathcal{D} corresponds to the points at infinity $\xi = \pm\infty$. By this coordinate transformation (cf., also, Balazs and Voros (1986), p. 117-118), the metric (49) is transformed into

$$ds^2 = \frac{4(dr^2 + r^2 d\phi^2)}{(1 - r^2)^2} = \frac{4(dx^2 + dy^2)}{(1 - x^2 - y^2)^2} \equiv \frac{4|dz|^2}{(1 - |z|^2)^2} \quad (53)$$

in the interior of the unit disc.

To summarize shortly: The coordinate transformations in (43) have, in the region sufficiently near the space-time singularity, accomplished to transform the Hamiltonian dynamics (originally formulated in the $(\Omega, \beta_+, \beta_-)$ variables and their canonical conjugate momenta) into that of piecewise “geodesic motion” (free motion) which takes place on the Poincaré unit disc (with constant negative curvature) inside a scattering domain $\mathcal{B} \subset \mathcal{D}$ (bounded by three walls) in the interior of the unit disc \mathcal{D} .

If one translates - via the defining coordinate transformations (51) - the equations (45) of the boundary of the scattering domain \mathcal{B} , one finds that they describe an equilateral triangle of geodesics with the three corners at “infinity” (at the boundary circle of the Poincaré disc). $\mathcal{B} \subset \mathcal{D}$ is thus a zero angle triangle on the Poincaré disc with finite hyperbolic area.

Using the identity $\tanh \xi \equiv 2 \tanh(\frac{\xi}{2})/(1 + \tanh^2(\frac{\xi}{2}))$, we express the formula for the left wall in the (x, y) coordinates. We get $4x/(1 + (x^2 + y^2)) = -1$, which is a circle $(x + 2)^2 + y^2 = 3$ with center $(x, y) = (-2, 0)$ and radius $r = \sqrt{3}$. The part of this circle which overlaps with the Poincaré disc \mathcal{D} governs the equation of the *left* wall $\partial\mathcal{B}_1$ ($2\pi/3 \leq \phi \leq 4\pi/3$). It is easily verified that $\partial\mathcal{B}_1$ cuts the boundary $\partial\mathcal{D} = \{(x, y) \mid x^2 + y^2 = 1\}$ of the Poincaré disc at the points $z = -\frac{1}{2} \pm i\frac{\sqrt{3}}{2} = \left\{ e^{i\frac{2\pi}{3}}, e^{i\frac{4\pi}{3}} \right\}$ and at a right angle $\varphi = \pi/2$. Hence, we conclude that the left boundary wall $\partial\mathcal{B}_1$ is a geodesic curve on the Poincaré unit disc \mathcal{D} . The triangular symmetry of the dynamical system gives us two similar ($2\pi/3$ rotated) wall boundaries consisting of circular arcs, the radii being $\sqrt{3}$ and centres located at $z = -2e^{\pm i2\pi/3} = 1 \pm i\sqrt{3}$. A *polygon* in the Poincaré disc \mathcal{D} is (as in Euclidian geometry) defined as a closed, connected area whose boundary consists in parts of geodesics, called its sides. Thus, the scattering domain \mathcal{B} is a regular *triangular polygon*; a regular 3-sided billiard (the three sides which we denote $\partial\mathcal{B}_1$, $\partial\mathcal{B}_2$ and $\partial\mathcal{B}_3$) on the Poincaré disc. (Cf. fig.9).

The periods of geodesic motion (free motion) which occur between the bounces against one of the three walls of the triangle correspond to the periods of straight line behavior (“Kasner epochs”) in fig.5. The source of the chaos of the model is to be thought of as due to the bounces in the potential boundaries $\partial\mathcal{B}$ in conjunction with the negative curved interior (which gives sensitive dependence on initial conditions. To the bounces are associated algebraic transition rules related to continued fraction

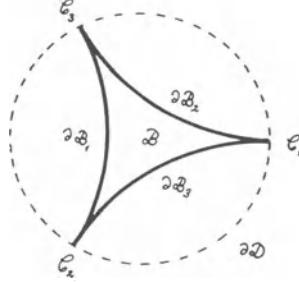


Figure 9. The triangular billiard $\mathcal{B} \subset \mathcal{D}$ forming the scattering domain inside the Poincaré disc $\partial\mathcal{D}$. The “billiard table” is bounded by the “infinitely hard” wall boundaries $\partial\mathcal{B} = \partial\mathcal{B}_1 \cup \partial\mathcal{B}_2 \cup \partial\mathcal{B}_3$. (Each side of the 3-sided polygon, $\partial\mathcal{B}_1$, $\partial\mathcal{B}_2$ and $\partial\mathcal{B}_3$, are geodesics on the Poincaré disc, being circular arcs orthogonal to the boundary circle $\partial\mathcal{D}$ (i.e. cutting the Poincaré disc boundary $\partial\mathcal{D}$ at right angles $\varphi = \pi/2$). Hence, \mathcal{B} is a triangular polygon, where all the interior angles between two sides at each of the three vertices C_1, C_2, C_3 are zero. The *approximations* involved in arriving at the description of the gravitational toy-model collapse are *very accurate* when sufficiently near the space-time “big crunch” singularity of the metric; The scattering potential $\tilde{V}(x, y)$ then (to a very high accuracy) vanishes identically in the interior of the domain \mathcal{B} (inside the wall boundaries $\partial\mathcal{B}$) and equals $+\infty$ outside \mathcal{B} .

expansions, cf. table 1. In fact, due to the symmetry properties of the domain (tiling the Poincaré disc by the action of an arithmetic group (E.B. Bogomolny *et al* (1992), J. Bolte *et al* (1992)), the chaos in the mixmaster collapse (in the region sufficiently near the singularity) is an example of arithmetic (algebraic) chaos. We thus appear to have an example of chaos of remarkable beauty.

Are the exact numerical values of the Lyapunov exponents λ an “artifact” of the considered set of transformations?

One observes, cf. J. Pullin (1990), that the constant negative curvature $K < 0$ of the interior of the Poincaré disc may be translated into a statement about a positive Lyapunov exponent. The geodesic motion takes place at constant velocity and we have for the Lyapunov exponent $\lambda = \{\text{velocity}\} \times \sqrt{-K} > 0$. (cf. Balazs and Voros (1986), p. 147). This expression scales with $\sqrt{-K}$ and the precise value of the negative curvature K of the Poincaré disc appears to be an artifact⁷⁰ of the selected set of coordinate transformations (43). However, the instability of a closed orbit γ in the triangular billiard on the Poincaré disc is given by multiplying the apparent “Lyapunov exponent” $\lambda \propto \sqrt{-K}$ with the hyperbolic length of the closed orbit. Since the length of a closed orbit $\oint_\gamma ds$ scales as $|K|^{-1/2}$ the expression for the instability exponents of the closed orbits is *invariant* under rescalings of K :

$$\{\text{instability of closed orbit}\} \propto \sqrt{-K} \times \{\text{length of closed orbit}\} =$$

⁷⁰One could transform the collapse orbit to a Poincaré disc with any negative curvature $K < 0$ (One should, however, verify how the velocity of the “billiard ball” scales in that case).

$$= \sqrt{-K} \oint_{\gamma} ds = \sqrt{-K} \oint_{\gamma} \frac{2|dz|}{\sqrt{-K}(1 - |z|^2)} = \oint_{\gamma} \frac{2|dz|}{(1 - |z|^2)} .$$

Thus the instability properties of the periodic orbit structure seem to be a (slightly) more invariant characterization.

As concerns the Wheeler-DeWitt toy-quantization of our gravitational collapse in the (t, ξ, ϕ) variables, say, it will actually have a complicated form (cf. equation (74) in C.W. Misner (1972)). Being sufficiently near the singularity of the metric where the walls are approximately infinitely hard we have an operator constraint on the wave function $\Psi = \Psi^{(3)}g = \Psi(t, \xi, \psi)$ of the form

$$\left\{ e^{-t} \frac{\partial}{\partial t} (-e^t \frac{\partial}{\partial t}) + \frac{1}{\sinh \xi} \frac{\partial}{\partial \xi} (\sinh \xi \frac{\partial}{\partial \xi}) + \frac{1}{\sinh^2 \xi} \frac{\partial^2}{\partial \phi^2} \right\} \Psi(t, \xi, \phi) = 0 \quad (54)$$

inside the domain $\mathcal{B} \subset \mathcal{D}$ with the boundary condition that Ψ vanishes at the boundary $\partial\mathcal{B}$.

The substitution

$$\Psi = \Psi(t, \xi, \phi) = e^{-(1/2+i\omega)t} \psi(\xi, \phi) \quad (55)$$

yields for the ψ component

$$\left\{ \frac{1}{\sinh \xi} \frac{\partial}{\partial \xi} (\sinh \xi \frac{\partial}{\partial \xi}) + \frac{1}{\sinh^2 \xi} \frac{\partial^2}{\partial \phi^2} \right\} \psi(\xi, \phi) = (\omega^2 + \frac{1}{4}) \psi(\xi, \phi) . \quad (56)$$

The left hand side is recognized as the Laplace operator on the two-dimensional Poincaré disc of constant negative curvature with metric $ds^2 = d\xi^2 + \sinh^2 d\phi^2$. We conclude that properties of solutions to the Wheeler DeWitt equation for the mixmaster collapse are one-to-one related to spectral properties of the Laplace operator on the Poincaré disc, that is the eigenvalue equation which in the more familiar Poincaré disc variables (x, y) reads

$$\Delta \psi = \frac{1}{4} (1 - x^2 - y^2)^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \psi = (\omega^2 + \frac{1}{4}) \psi . \quad (57)$$

with the boundary condition that ψ vanishes at the boundary $\partial\mathcal{B}$ of the triangular domain.

Moreover, spectral properties of the Laplace operator are related to the Gutzwiller trace formula involving a sum over the periodic orbits inside the triangular domain $\mathcal{B} \subset \mathcal{D}$. (Cf. Gutzwiller (1990) and also Cvitanovic (1990) and references therein).

It is amusing that the spectrum of the Laplace operator exhibits “non-generic” energy-level statistics in the sense that the distributions of nearest neighbor level spacings displays Poisson rather than Wigner statistics. This is due to the symmetry properties of our triangular domain (the domain tiles the disc under the action of a so-called “*arithmetic group*”) with a resulting exponentially large *degeneracy* of lengths of the periodic orbits. Note, also, that the (approximate) semiclassical Gutzwiller trace formula, cf. Gutzwiller (1971, 1990), is exact in this case, since it coincides with the exact Selberg trace formula (giving an exact relationship between the quantum spectrum and the classical periodic orbits).⁷¹

In the context of the toy-model mixmaster gravitational collapse, a contact is thus made between the disciplines of “Quantum Cosmology” and “Quantum Chaos”, each dating back twenty years to pioneering works of C.W. Misner and M.C. Gutzwiller!

⁷¹For a discussion of the most recent results as concerns the the Wheeler-DeWitt quantization of the mixmaster gravitational collapse, see R. Graham (1993).

6. DISCUSSION, OPEN PROBLEMS AND DREAMS (SOME FINAL REMARKS)

The toy-model sketched (the mixmaster gravitational collapse) is a very simple model. Nevertheless, it has enough degrees of freedom to behave in a dynamically complicated and non-predictable way (with positive Kolmogorov entropy) on approach to the singular space-time point (the “big crunch” point) and in this way the toy-model space-time metric captures an interesting non-linear aspect of the classical Einstein equations (probing the Einstein equations in a domain where the field strength gets strong (as captured by an invariant like $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$, which grows without limit) and non-linearities of the Einstein equations are important).

In fact, if one were to imagine a *generic* gravitational collapse (e.g. for an entire “model-universe”) to a singular point, then it would misguide our intuition to imagine a Friedmann Robertson Walker type of collapse solution (the “standard” big crunch/big bang). This (non-generic) solution is completely integrable (and behaves in a dynamically trivial way) as an artifact of the enormous amount of symmetry imposed on that solution.

Rather one should try to imagine a sort of “metrical chaos” (chaos in the space-time metric) being developed on approach to the curvature singularity - as the gravitational field strength gets stronger and non-linear aspects of the Einstein equations are increasingly important.

6.1. How to build up concepts which characterize such “metrical chaos”?

The general problem of transferring standard indicators of chaos to the situation of general relativity (pointed out in S.E.Rugh (1990a), emphasized in S.E. Rugh (1990b) and by J. Pullin (1990)) is well illustrated by the simple spatially homogeneous mixmaster collapse.

We have noticed (S.E. Rugh (1990 a,b)) that a naive application of a “standard” Lyapunov exponent will be a highly gauge-dependent measure, even under the small subclass of gauge-transformations, which only involve (the perfectly allowed) reparametrizations of the time-coordinate. To construct a gauge invariant generalization of a “Lyapunov exponent” is a yet unsolved problem.

Perhaps, there is a “no go” theorem to be constructed towards a completely gauge invariant construction? (H.B. Nielsen and S.E. Rugh)

We have noticed that the ergodicity properties of the toy-model mixmaster collapse are interesting (and subtle): In some “gauges” the phase space is ever expanding (or ever collapsing) and there are no return properties of orbits and a concept like “ergodicity”, for example, is very ill defined. By a suitable set of coordinate transformations, however, and some very reliable approximations, the Hamiltonian can be brought into that of a geodesic flow on a two-dimensional Riemann manifold of constant negative curvature (the Poincaré disc) bounded by some stationary walls forming an equilateral triangle with three corner infinities on the Poincaré disc. In this picture a number of stochastic properties of the gravitational collapse may be stated, e.g., the property of ergodicity. It is even a K-flow!

However, the chaotic properties of a toy-model like the mixmaster gravitational collapse is of course - inherently - the same, irrespective of the choice

of “gauge” (choice of description). If we get apparent “chaotic behavior” in one description (in one gauge) but absence of chaotic behavior in another description - it is because one has posed the wrong question!

I thus disagree with the conclusion arrived at by Pullin (cf. also J. Pullin (1990) about the superiority of the “Poincaré disc” gauge-choice over other gauge-choices.

It appears to me that no “gauge” is better than others. Rather, one ought to seek measures (of chaos) which are invariant under (some large class of) gauge transformations (to the extent they are possible to construct).

6.2. Is chaos in spacetime metrics of importance for observable phenomena in Nature?

The extreme conditions for the gravitational field referred to in this manuscript (probing the Einstein equations near space-time curvature singularities) are of course far removed from “daily life” gravity. It is hard to find observable *astrophysical* phenomena which involve very strong gravitational fields and thus probe the *non-linearities* of the Einstein equations at a really deep level. (For instance, the rather large perihelion precession in binary pulsar systems, which are indeed marvellous relativistic laboratories, do not qualify as an observational test of non-linearities of the Einstein equations at a deep level). In the description of many gravitational phenomena, even the (quasi) *linearized* equations suffice.

The relevant laboratory for probing deep non-linearities of the Einstein equations seems to be (1) violent astrophysical phenomena ⁷² (2) early moments of the Universe (could the Universe itself - and the large scale structure of the spacetime metric of the Universe - have had a violent past? Would there be observable fingerprints of this today? Cf. also next subsection 6.3)

Turning to our toy-model metric (3) the Einstein equations for the metric involve exponential functions. Such exponential functions are well known to give out “large numbers” very fast! In fact, it is so that if we - by hand - put some “reasonable scale” (like $L \sim c/H \sim 10^{28} \text{ cm} \sim$ the characteristic size of the present Universe) for the initial length scales, $a \sim b \sim c \sim 10^{-28} \text{ cm}$ of the gravitational collapse, then only few (chaotic) cycles of the the scale-functions a, b, c occur before they reach and pass the Planck regime, $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta} \sim l_{Pl}^{-4}$. Thus, the gravitational collapse is chaotic (“metrical chaos”) in the very same region where it should be treated quantum mechanically. In that sense this gravitational collapse is very much addressing a topic on the interface of quantum physics, cosmology and chaos.

If the large scale structure of spacetime (our Universe) started out (near Planck time, say) in a chaotic, oscillatory behavior given by the mixmaster metric (3) say, would we then be able to detect signals of this behavior today? Most likely, we would observe a substantially larger anisotropy of the background microwave radiation temperature than the high degree of isotropy observed today, since the Universe dissipates away the initial anisotropies too slowly. (See also next subsection). In a Universe with an *inflationary* phase, the existence of anisotropies and inhomogeneities before the onset of inflation is not excluded, but it appears that there is not much time for spacetime dynamics to

⁷²A gravitational phenomenon like the generation of gravitational waves (which will be observable in the next century) emitted from *strong* astrophysical sources needs that we take full account of the nonlinearity of the Einstein equations, cf. e.g. D. Christodoulou (1991).

perform “chaotic oscillations” in the short interval from Planck times $t_{Pl} = \sqrt{(\hbar G/c^5)} \sim 10^{-43}$ seconds after the “big bang” to the onset of an inflationary phase at the G.U.T. scales $\sim 10^{14}$ GeV occurring $\sim 10^{-34}$ seconds after the “big bang”.⁷³ Moreover, inflation tends to smooth things out so pre-inflationary history (e.g. dynamics of the spacetime metric) is rendered more irrelevant.

A remark concerning dependence of the “present” on the “past”. I.

A more serious problem than whether we will be able to detect evidence of “chaotic spacetimes” in the early moments of the Universe is whether detectable signals (now) in cosmology contain information which will prove useful in probing the structure of the Natural laws at higher energies than those which can be created here on Earth. In view of the foreseeable limitations in our abilities of putting substantially more than present energies $\sim 100\text{GeV}$ on a single elementary particle (experiments here on Earth), one would like to “resort to” our Universe - in its very early moments - as a laboratory to probe the structure of fundamental laws at energies beyond $\sim 100\text{GeV}$. In cosmology, however, we really see the big bang physics through an extremely cloudy and *little informative filter!* See also, e.g., H.B. Nielsen, S.E. Rugh and C. Surlykke (1993). Every time *thermal equilibrium phases* are reached at certain stages in the evolution of the Universe then only very little information of the physics that went on *before* that phase can reach us today. It is difficult for a signal to survive through an equilibrium phase! Basically, only *conserved* quantum numbers *survive*, like baryon and lepton numbers and energy. However, if the thermodynamic equilibrium is not reached globally, there may survive some information in the correlations or, rather, in the *spatial variations* in the densities of these conserved quantities. Gravitational signals, i.e. spacetime dynamics, may contribute to establish correlations over big distances, since gravity is a long range force.

6.3. A chaotic initial “big bang”?

Could our Universe have started out in a chaotic state? Cf. the “chaotic cosmology” concept, developed by C.W. Misner *et al* , which attempts at creating our present Universe from (almost) arbitrary irregular initial conditions, e.g. from a spacetime metric with large initial anisotropies and inhomogeneities.⁷⁴

A remark concerning dependence of the “present” on the “past”. II.

Whereas one would (in cosmology) be tempted to explain the present Universe without having to resort to considerable fine tunings of the initial data of the Universe (at Planck scales, say) a “high energy physicist” would prefer that the present depends very sensitively on the past!⁷⁵ - in order to be able to deduce interesting *information about the past* (in particular, to get a detailed look into the Natural laws beyond $\sim 100\text{GeV}$, via signals from cosmology).

⁷³Using $R \propto T^{-1}$ (valid after inflation has occurred), it is fascinating to imagine that our current Hubble volume (size of the Universe) had a size of about ~ 10 cm at those early times! (i.e. when the temperature in the Universe was $\sim 10^{14}$ GeV).

⁷⁴See also, e.g., discussions in B.L. Hu *et al* (1993).

⁷⁵Inflation, which only lessens the dependence of the present from the initial data of the past, is in that perspective not good news. (Even within the inflationary Universe, however, there will be an arbitrary large number of initial conditions which are inconsistent with observations today).

The viewpoint depends on the question(s) which drives our study: Do we “want” to give a “probabilistic” explanation of the present observed Universe (I.e.: if the Universe we see today results from a larger basin of initial conditions then it is more “likely” to have it?) or do we “want” signals coming from cosmology to be a very sharp probe exploring into the structure of Natural laws at very high energies (corresponding to hot dense phases of the Universe)?

If one contemplates, cf. e.g. S. Weinberg (1992), that the actually implemented Natural laws are the only *self-consistent* set of laws and parameters imaginable, and also consider initial conditions of the Universe as resulting from such implemented regularities (the division line between Natural laws and initial conditions is rather arbitrary anyway) then I must admit that I have no problem with reconciling myself to the fact that *initial conditions* of the Universe may be as fine tuned as *Natural laws* appear to be (so far explored).

If the Universe had very large irregularities initially, we have - in order to obtain the *present* Universe with large scale smoothness - to introduce some physical mechanisms to “dissipate away” all these initial irregularities (anisotropies and inhomogeneities).

A very effective dissipative process, suggested by Zel'dovich, is that of vacuum particle creation from the changing gravitational field in anisotropic spacetimes. (Gravitational fields can create particles somewhat similar to the way magnetic fields can create electron pairs). In a test field approximation where the spacetime anisotropies are not too big, one may estimate the probability \mathcal{P} of producing a pair of conformally invariant scalar particles in a world tube T of constant, co-moving cross section extending from the singularity to the present and it turns out to be proportional to the integrated Weyl-curvature invariant over the world tube (cf., e.g., J.B. Hartle (1981)),

$$\mathcal{P} \sim \int_T d^4x \sqrt{-g} C_{\alpha\beta\gamma\delta} C^{\alpha\beta\gamma\delta} .$$

If the initial Weyl tensor invariant is too large and if the Universe has managed to dissipate away all this initial anisotropy, the heat entropy generated would be very large, and the Universe would *already* have reached “heat death” which is not the case. In fact, the “small” entropy per baryon now severely limits the amount of dissipation that has taken place in the initial phases of the Universe⁷⁶

In order not to reach “heat death” too fast, the Universe has to start out in a state of low entropy⁷⁷. According to the Weyl tensor hypothesis of R. Penrose (1979), gravitational entropy should, somehow, be measured by the Weyl curvature tensor, and the constraint imposed on the Universe to start out in a state of low entropy translates, e.g., into a requirement of creating a Universe with a small Weyl curvature (whereas a final “big crunch”, if the Universe is closed, may - and is expected to - have a large Weyl curvature and a large degree of disorder).

The mixmaster space-time metric appears to be a well suited toy-model for implementation of the Weyl tensor hypothesis, having a very large Weyl curvature near the space-time singularity.⁷⁸ Thus, according to the Weyl-curvature hypothesis, the chaotic

⁷⁶In order not to over-produce entropy. See also Barrow and Matzner (1977) and Barrow (1978)).

⁷⁷This fact, and the enormous amount of fine tuning and precision in the organization of the starting conditions of the Universe this implies, has been emphasized very strongly in e.g. R. Penrose (1989). Perhaps, the Guth/Linde inflationary Universe provides us, more naturally, with a “low entropy” initial condition? (see, e.g., D. Goldwirth and T. Piran (1991) and references therein).

⁷⁸M. Biesiada and S.E. Rughe, in preparation.

collapse should be considered a (simplified) candidate for the large scale metric at a final crunch⁷⁹ rather than at the initial bang.

Turning back to our question about observational consequences of having chaos in spacetime metrics: If it is a chaotic collapse at a “big crunch” point - it will probably be too hot for us to be there and observe it!

ACKNOWLEDGMENTS

I would like to thank Holger Bech Nielsen for all sorts of fruitful exchanges, from chaos in dynamical systems to “chaos in Natural laws” to viewpoints concerning what our Creator should or should not do.

I am indebted to the “Chaos Group” at the Niels Bohr Institute and NORDITA and John Barrow, Freddy Christiansen, Marek Demianski, Robert Graham, Bernard Jones, V.N. Lukash and Igor Novikov and, especially, I am indebted to Holger Bech Nielsen and Hans Henrik Rugh for discussions and collaboration on various aspects of the gravitational collapse.

Special thanks are also due to Marek Biesiada for discussions and collaboration in Copenhagen, Poland and Japan and hospitality at the Copernicus Astronomical Center at Warsaw, Poland.

Support from the Danish Natural Science Research Council Grant No. 11-8705-1 is gratefully acknowledged and I thank Dave Hobill for the invitation to participate in this interesting workshop.

REFERENCES

- Arnowitt, R., Deser, S. and Misner, C.W., 1962, in *Gravitation: An introduction to current research*, edited by L. Witten. Wiley. New York.
- Artuso, R., Aurell, E. and Cvitanovic, P., 1990, Recycling of strange sets: II. Applications, *Nonlinearity* **3**, 361.
- Ashtekhar, A. and Pullin, J., 1989, Bianchi Cosmologies: A New Description, Syracuse University, print **89-0873**, Nathan Rosen Festschrift (Israel Physical Society, 1990).
- Balazs, N.L. and Voros, A., 1986, Chaos on the pseudosphere, *Phys. Rep.* **143**, 109.
- Barrow, J.D., 1978, Quiescent cosmology, *Nature* **272**, 211-215.
- , 1982, Chaotic Behavior in General Relativity, *Phys. Rep.* **85**, No. 1, 1-49.
- , 1984, in *Classical General Relativity*, edited by W.B. Bonnor et al., Cambridge University Press.
- Barrow, J.D. and Cotsakis, S., 1989, Chaotic behavior in higher-order gravity theories, *Phys. Lett* **232 B**, 172-176.
- Barrow, J.D. and Matzner, R.A., 1977, The homogeneity and isotropy of the Universe, *Mon. Not. R. astr. Soc.* **181**, 719-727.
- Barrow, J.D. and Silk, J., 1984, *The Left Hand of Creation*. Heinemann, London.
- Barrow, J.D. and Sirouse-Zia, M., 1989, Mixmaster cosmological model in theories of gravitation with a quadratic Lagrangian, *Phys. Rev. D* **39**, 2187.

⁷⁹Cf. also the title “Gravity’s chaotic future” in a recent issue of *Science News*, Vol. 144, p. 369-384. Dec.4 (1993).

- Bedford, T., Keane, M. and Series, C. (eds), 1991, *Ergodic theory, symbolic dynamics and hyperbolic spaces*. Oxford University Press.
- Bell, J.S., 1987, Speakable and unspeakable in quantum mechanics, Cambridge University Press.
- Berry, M., 1983, Semiclassical mechanics of regular and irregular motion, in Les Houches, Session XXXVI, 1981, edited by G.Ioos et al. North-Holland.
- , M., 1987, Quantum Chaology, *Proc. Roy. Soc. Lond. A* **413**, 183.
- Behr, C., 1962, Eine Verallgemeinerung des Friedmannschen Weltmodells mit positiver Raumkrümmung, *Zeitschrift für Astrophysik* **54**, 268.
- Belinskii, V.A. and Khalatnikov, I.M., 1969, On the Nature of the Singularities in the General Solution of the Gravitational Equations, *Sov. Phys. JETP* **29**, No. 5, 911-917.
- Belinskii, V.A., Khalatnikov, I.M. and Lifshitz, E.M., 1970, Oscillatory Approach to the Singular Point in the Relativistic Cosmology, *Adv. Phys.* **19**, 525-573.
- , 1982, A general solution of the Einstein equations with a time singularity, *Adv. Phys.* **31**, 639-667.
- Berger, B.K., 1989, Quantum chaos in the mixmaster universe, *Phys. Rev. D* **39**, 2426-2429.
- , 1990, Numerical study of initially expanding mixmaster universes, *Class. Quant. Grav.* **7**, 203.
- , 1991, Comments on the Computation of Liapunov Exponents for the Mixmaster Universe, *Gen. Rel. Grav.* **23**, 1385.
- , 1993, How to determine approximate Mixmaster parameters from numerical evolution of Einstein's equations, preprint.
- Biesiada, M. and Rugh, S.E., 1994, Maupertuis principle, Wheelers superspace and chaotic gravitational collapses, to appear.
- Biró, T.S., Gong, C., Müller, B. and Trayanov, A., 1993, Hamiltonian Dynamics of Yang-Mills Fields on a Lattice, Duke University preprint.
- Bogoyavlensky, O.I. and Novikov, S.P., 1973, Singularities of the cosmological model of the Bianchi IX type according to the qualitative theory of differential equations, *Sov. Phys. JETP* **37**, 747.
- Bogoyavlensky, O.I., 1985, *Methods in the Qualitative Theory of Dynamical Systems in Astrophysics and Gas Dynamics*, Springer Verlag, New York.
- Bogomolny, E.B., Georgeot, B., J.-Giannoni, M. and Schmit, C., 1992, Chaotic Billiards Generated by Arithmetic Groups, *Phys. Rev. Lett.* **69**, 1477.
- Bohr, H. and Nielsen, H.B., 1990, Limits where chaos tends to order and complexity to regularity. In *Characterizing Complex Systems*, Ed. H. Bohr. World Scientific.
- Bohr, T., 1990, Chaos and Turbulence, in *Applications of Statistical Mechanics and Field Theory to Condensed Matter*, edited by A.R. Bishop et al. Plenum Press.
- Bolte, J., Steil, G. and Steiner, F., 1992, Arithmetical Chaos and Violation of Universality in Energy Level Statistics, DESY **92-061** preprint.
- Bombelli, L. and Calzetta, E., 1992, Chaos around a black hole, *Class. Quant. Grav.* **9**, 2573.
- Bugalho, M.H., Rica da Silva, A. and Sousa Ramos, J., 1986, The Order of Chaos on A Bianchi-IX Cosmological Model, *Gen. Rel. Grav.* **18**, 1263.
- Burd, A., Buric, N. and Ellis, G.F.R., 1990, A Numerical Analysis of Chaotic Behavior in Bianchi IX Models, *Gen. Rel. Grav.* **22**, No.3, 349.

- Burd, A., Buric, N. and Tavakol, R.K., 1991, Chaos, entropy and cosmology, *Class. and Quant. Gravity* **8**, 123.
- Burd, A. and Tavakol, R.K., 1993, Invariant Lyapunov exponents and chaos in cosmology, *Phys. Rev. D* **47**, 5336.
- Chang, S.J. and Weiss, N., 1979, Instability of constant Yang-Mills fields, *Phys. Rev. D* **20**, 869.
- Chernoff, D.F. and Barrow, J.D., 1983, Chaos in the Mixmaster Universe, *Phys. Rev. Lett* **50**, 134.
- Christodoulou, D., 1991, Nonlinear Nature of Gravitation and Gravitational-Wave Experiments, *Phys. Rev. Lett* **67**, 1486.
- Coleman, S., Hartle, J.B., Piran, T. and Weinberg, S. (eds), 1992, *Quantum Cosmology and Baby Universes*. World Scientific.
- Coley, A.A. and Tavakol, R.K., 1992, Fragility in Cosmology, *Gen. Rel. Grav.* **24**, 835.
- Contopoulos, G., 1990, Periodic orbits and chaos around two black holes, *Proc. Roy. Soc. London A* **431**, p. 183-202.
- Contopoulos, G., Grammaticos, B. and Ramani, A., 1993, Painlevé analysis for the mixmaster universe model, *J. Phys. A: Math. Gen.* **26**, 5795.
- Corless, R.M., Frank, G.W. and Monroe, J.G., 1990, Chaos and Continued Fractions, *Physica D* **46**, 241-253.
- Cotsakis, S., Demarett, J., De Rop, Y. and Querella, L., 1993, Mixmaster Universe in fourth-order gravity theories, *Phys. Rev. D* **48**, 4595.
- Cvitanovic, P., 1989, *Universality in Chaos*, 2nd. edition. Adam Hilger.
- Cvitanovic, P., 1992, The Power of Chaos, in *Applications of Chaos, Electrical Power Research Institute Workshop 1990*, edited by J.H. Kim and J. Stringer. John Wiley & Sons. New York. In particular, chapt. 3, Cycles as the skeleton of chaos.
- Cvitanovic, P. and Myrheim, J., 1989, Complex Universality, *Comm. Math. Phys.* **121**, 225.
- Eathetal D'Eath, P.D., Hawking, S.W. and Obregón, O., 1993, Supersymmetric Bianchi models and the square root of the Wheeler-DeWitt equation, *Phys. Lett B* **300**, 44.
- Demarett, J and De Rop, Y., 1993, The fractal nature of the power spectrum as an indicator of chaos in the Bianchi IX cosmological model, *Phys. Lett B* **299**, 223.
- DeWitt, B.S., 1967, Quantum Theory of Gravity. I. The Canonical Theory, *Phys. Rev.* **160**, 1113.
- DeWitt, B.S., 1970, Spacetime as a sheaf of geodesics in superspace, in *Relativity*, edited by M. Carmeli *et al*. Plenum Press. New York.
- Dirac, P.A.M., 1959, Fixation of Coordinates in the Hamiltonian Theory of Gravitation, *Phys. Rev* **144**, 924.
- Doroshkevich, A.G. and Novikov, I.D., 1971, Mixmaster Universes and the Cosmological Problem, *Sov. Astron.-AJ* **14**, 763.
- Doroshkevich, A.G., Lukash, V.N. and Novikov, I.D., 1973, The isotropization of homogeneous cosmological models, *Sov.Phys.-JETP* **37**, 739.
- Eckmann, J.P. and Ruelle, D., 1985, Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.* **57**, 617.
- Feigenbaum, M.J., 1988, Presentation Functions, Fixed Points, and a Theory of Scaling Function Dynamics, *Journ. Stat. Phys.* **52**, 527.
- Francisco, G. and Matsas, G.E.A., 1988, Qualitative and Numerical Study of Bianchi IX Models, *Gen. Rel. Grav.* **20**, No. 10, 1047.

- Froggatt, C.D. and Nielsen, H.B., 1991, *Origin of Symmetries*, World Scientific.
- Furusawa, T., 1986, Quantum Chaos in the Mixmaster Universe, *Prog. Theor. Phys.* **75**, 59-67 and **76**, 67-74.
- Giannoni, M.-J. and Ullmo, D., 1990, Coding Chaotic Billiards, *Physica D* **41**, 371.
- Gibbons, 1985, Solitons in General Relativity and Supergravity, in *Nonlinear Phenomena in Physics*, edited by F. Claro, pp.255-269. Springer Verlag.
- Gibbons Gibbons, G.W., 1990, Self-gravitating magnetic monopoles, global monopoles and black holes, D.A.M.T.P. preprint (1990), to appear in the Proceedings of XII Autumn School, Lisbon 1990 (World Scientific).
- Goldwirth, D.S. and Piran, T., 1991, Entropy, Inflation and the arrow of time, *Class. Quant. Grav* **8**, L 155.
- Graham, R., 1991, Supersymmetric Bianchi Type IX Cosmology, *Phys. Rev. Lett* **67**, 1381.
- , 1992, Supersymmetric general Bianchi type IX cosmology with a cosmological term, *Phys. Lett B* **277**, 393.
- , 1993, Chaos and Quantum Chaos in Cosmological Models, preprint.
- Graham, R. and Szépfalusy, P., 1990, Quantum creation of a generic universe, *Phys. Rev. D* **42**, 2483-2490.
- Grassberger, P., 1986, Toward a Quantitative Theory of Self-Generated Complexity, *Int. Journ. Theor. Phys.* **25**, 907.
- Grassberger, P., Schreiber, T. and Schaffrath, C., 1991, Nonlinear Time Sequence Analysis, *Int. Journ. of Bifurcation and Chaos* **1**, 521.
- Guth, A.H., 1992, The Big Bang and Cosmic Inflation, MIT preprint CTP # 2144, to appear in *The Oscar Klein Memorial Lectures. Volume 2*, edited by G. Ekspong. World Scientific.
- Gutzwiller, M.C., 1971, Periodic Orbits and Classical Quantization Conditions, *Jour. Math. Phys.* **12**, 343.
- , 1990, *Chaos in Classical and Quantum Mechanics*. Springer Verlag. New York.
- Hardy, G.H. and Wright, E.M. 1938, *The Theory of Numbers*. Oxford.
- Hartle, J.B., 1981, Particle Production and Dynamics in the early Universe, in *Quantum Gravity 2. A second Oxford symposium*, edited by C.J. Isham, R. Penrose and D.W. Sciama. Clarendon Press. Oxford.
- Hartle, J.B. and Hawking, S.W., 1983, Wave function of the Universe, *Phys. Rev D* **28**, 2960.
- Hartle, J.B. and Kuchar, K.V., 1984, The Role of Time in Path Integral Formulations of Parametrized Theories, pp. 315-326 in *Quantum Theory of Gravity. Essays in honor of the 60th birthday of Bryce S. DeWitt*, edited by S.M. Christensen. Adam Hilger.
- Hawking, S.W. and Ellis, G.F.R., 1973, *The large scale structure of space-time*. Cambridge University Press.
- Hawking, S.W. and Luttrell, J.C., 1984, The isotropy of the Universe, *Phys. Lett. B* **143**, 83-86.
- Hawking, S.W. and Penrose, R., 1970, The singularities of gravitational collapse and cosmology, *Proc. Roy. Soc. Lond. A* **314**, 529.
- Hobill, D., Bernstein, D., Simpkins, D. and Welge, M., 1989, How chaotic is the mixmaster universe?, *Proc. 12th Int. Conf. Gen. Rel. Grav. (University of Colorado, Boulder)*, abstracts p. 337.
- Hobill, D., Bernstein, D., Welge, M. and Simpkins, D., 1991, The Mixmaster cosmology as a dynamical system, *Class. Quant. Grav.* **8**, 1155-1171.

- Hu, B.L., Paz, J.P. and Sinha, S., 1993, pp. 145-165 in *Directions in General Relativity. Volume 1. Papers in honor of Charles Misner*, edited by B.L. Hu, M.P. Ryan and C.V. Vishveshwara. Cambridge University Press.
- Karas, V. and Vokrouhlický, D., Chaotic Motion of Test Particles in the Ernst Space-time, 1992, *Gen. Rel. Grav.* **24**, 729.
- Kaufmann, Z. and P. Szépfalusy, 1989, Properties of the entropies at weak intermittent states of Lorentz-type systems, *Phys. Rev. A* **40**, 2615.
- Khalatnikov, I.M., Lifshitz, E.M., Khanin, K.M., Shur, L.N. and Sinai, Ya.G., 1985, On the Stochasticity in Relativistic Cosmology, *J. of Stat. Phys.* **38**, No 1/2, 97-114 and references therein.
- King, D.H., 1991, Gravity-wave insights to Bianchi type-IX universes, *Phys. Rev. D* **44**, 2356.
- Kolb, E.W. and Turner, M.S., 1990, *The early universe*. Addison-Wesley.
- Kuchar, K.V. and Ryan, M.P., 1989, Is minisuperspace quantization valid? Taub in Mixmaster, *Phys. Rev. D* **40**, 3982.
- Landau, L.D. and Lifshitz, E.M., 1975, *The Classical Theory of Fields*. Pergamon (4.ed.), especially §116-119.
- Larsen, A.L., 1993, Chaotic string-capture by black hole, NORDITA preprint, NORDITA-93/55 P.
- Lichtenberg, A.J. and Lieberman, M.A., 1983, *Regular and Stochastic Motion*. Springer Verlag.
- Lin, X. and Wald, R.M., 1990, Proof of the closed-universe recollapse conjecture for general Bianchi type-IX cosmologies, *Phys. Rev. D* **41**, 2444.
- Lloyd, S. and Pagels, H., 1988, Complexity as Thermodynamic Depth, *Annals of Physics* **188**, 186.
- Lukash, V.N., 1983, Doctorate Thesis. Moscow.
- , 1989-90, Private communication.
- MacCallum, M.A.H., 1979, Anisotropic and Inhomogeneous Relativistic Cosmologies, in *General Relativity - An Einstein Centenary Survey*, edited by S.W. Hawking and W. Israel. Cambridge University Press.
- , 1983, Relativistic Cosmology for Astrophysicists, pp. 9-39 in Jones, B.J.T. Jones, J.E. (eds.), *The Origin and Evolution of Galaxies*, D. Reidel Publ.
- Matinyan, S.G., 1985, Dynamical chaos of non-Abelian gauge fields, *Sov.J. Part. Nucl.* **16** (3), 226.
- Meyer, H.-D., Theory of Liapunov exponents of Hamiltonian systems and a numerical study on the transition from regular to irregular classical motion, *J. Chem. Phys.* **84**, 3147.
- Misner, C.W., 1968, The Isotropy of the Universe, *Astrophys. J.* **151**, 431.
- , 1969 a, Mixmaster Universe, *Phys. Rev. Lett.* **22**, No. 20, 1071.
- , 1969 b, Quantum Cosmology. I. *Phys. Rev.* **186**, 1319.
- , 1969 c, Absolute Zero of Time, *Phys. Rev.* **186**, 1328.
- , 1972, Minisuperspace, in *Magic Without Magic: John Archibald Wheeler, a Collection of Essays in Honor of his 60 th Birthday*, edited by J. Klauder. Freeman. San Francisco.
- Misner, C.W., Thorne, K.S. and Wheeler, J.A., 1973, *Gravitation*. W.H.Freeman & Co, especially §30. Cited in the text as MTW.
- Moncrief, V. and Ryan, M.P., 1991, Amplitude-real-phase exact solutions for quantum mixmaster universes, *Phys. Rev. D* **44**, 2375.

- Müller, B. and Trayanov, A., 1992, Deterministic Chaos in Non-Abelian Lattice Gauge Theory, *Phys. Rev. Lett.* **68**, 3387.
- Nielsen, H.B., 1976, Dual Strings, Section 6. Catastrophe Theory Programme, pp. 528-543 in Barbour, I.M. and Davis, A.T. (eds.), *Fundamentals of Quark Models*, Scottish University Summer School in Physics.
- , 1981, Field Theories without fundamental gauge symmetries, *Phil. Roy. Soc. Lond. A* **320**, 261.
- Nielsen, H.B. and Rugh, S.E., 1992 a, Chaos in the Fundamental Forces?, Niels Bohr Institute preprint NBI-HE-92-85, appeared in *Quantum Physics and the Universe*, edited by M. Namiki et al., *Vistas in Astronomy* **37**. Pergamon Press.
- , 1992 b, Weyl Particles, Weak Interactions and Origin of Geometry, *Nucl. Phys. (Proc. Suppl.)* **29 B,C**, 200-246.
- , 1993, Why do we live in 3+1 dimensions?, Niels Bohr Inst. preprint, NBI-HE-93-11.
- Nielsen, H.B., Rugh, S.E. and Surlykke, C., 1993, Seeking inspiration from the Standard Model in order to go beyond it, Niels Bohr Inst. preprint, NBI-HE-93-48.
- Nielsen, N.K. and Olesen, P., 1978, An unstable Yang-Mills field mode, *Nucl. Phys. B* **144**, 376.
- Olive, K.A., 1990, Inflation, *Phys. Rep.* **190**, 307-403.
- Ove, R., 1990, Nonlinear Gravitational Effect, *Phys. Rev. Lett.* **64**, 1200.
- Penrose, R., 1979, Singularities and time-asymmetry. In *General Relativity - An Einstein Centenary Survey*, edited by S.W. Hawking and W. Israel. Cambridge University Press.
- , 1989, *The Emperor's New Mind*. Oxford University Press.
- Pullin, J., 1990, Time and Chaos in General Relativity. Talk given at the *VII SILARG, Simposio Latinoamericano de Relatividad y Gravitacion*, Mexico City, Mexico, December 1990. Syracuse University preprint **90-0734**.
- Qadir, A. and Wheeler, J.A., 1989, Late Stages of Crunch, *Nucl. Phys. (Proc. Suppl.)* **6 B**, 345-347.
- Raychaudhuri, A.K., 1955, Relativistic Cosmology. I, *Phys. Rev.* **98**, 1123.
- , 1989, An approach to Anisotropic Cosmologies, in *Gravitation, Gauge Theories and the Early Universe*, edited by B.R. Iyer et al., pp. 89-106. Kluwer.
- Ruelle, D., 1991, *Chance and Chaos*. Princeton University Press.
- Rugh, S.E., 1990a, *Chaotic Behavior and Oscillating Three-volumes in a Space-Time Metric in General Relativity*, Cand.Scient.Thesis, The Niels Bohr Institute, Copenhagen (January 1990; Second printing, with small additions and corrections, april 1990; distributed). Available upon request.
- , 1990b, Chaos in the Einstein Equations, Niels Bohr Institute preprint NBI-HE-91-59. Never submitted but distributed as a preliminary draft at the Texas/ESO-CERN symposium on Relativistic Astrophysics, Cosmology and Fundamental Physics, December 16-21 1990, Brighton, UK.
- , 1993, Chaos in the Einstein equations, to appear in *Chaos, Fractals and Solitons*.
- , 1994, Chaos in the Fundamental Forces - Characterization and Importance?, *Licentiate Thesis*, The Niels Bohr Institute, to appear.
- Rugh, S.E. and Jones, B.J.T., 1990, Chaotic behavior and oscillating three-volumes in Bianchi IX universes, *Phys. Lett. A* **147**, 353.
- Ryan, M.P., 1972, *Hamiltonian Cosmology*. Springer Verlag.

- Ryan, M.P. and Shepley, L.C., 1975, *Homogeneous Relativistic Cosmologies*. Princeton University Press.
- Savidy, G.K., 1984, Classical and quantum mechanics of non-Abelian gauge fields, *Nucl.Phys. B* **246**, 302.
- Smolin, L., 1990, New Hamiltonian variables, pp. 449-461 in *Proc. 12.th Int. Conf. on General Relativity and Gravitation*, Ashby, N. et al (eds.), Cambridge University Press.
- Straumann, N. and Zhou, Z., 1990, Instability of a colored black hole solution, *Phys. Lett. B* **243**, 33.
- Szépfalusy, P. and Györgyi, G., 1986, Entropy decay as a measure of stochasticity in chaotic systems, *Phys. Rev A* **33**, 2852.
- Szydłowski, M. and Lapeta, A., 1990, Pseudo-Riemannian manifold of mixmaster dynamical systems, *Phys. Lett B* **148**, 239.
- Szydłowski, M. and Biesiada, M., 1991, Chaos in mixmaster models, *Phys. Rev. D* **44**, 2369.
- Tavakol, R.K., 1991, Fragility and Deterministic Modelling in the Exact Sciences, *Brit.J.Phil.Sci.* **42**, 147.
- Taub, A.H., 1951, Empty spacetimes admitting a three parameter group of motions, *Ann. of Math.* **53**, 472.
- Thorne, K.S., 1985, The Dynamics of Spacetime Curvature: Nonlinear Aspects, in *Nonlinear Phenomena in Physics*, edited by F. Claro, pp.280-291. Springer Verlag.
- Wald, R.M., 1984, *General Relativity*. The University of Chicago Press.
- Weinberg, S., 1992, *Dreams of a Final Theory*. Pantheon Books. New York.
- Wellner, M., 1992, Evidence for a Yang-Mills Fractal, *Phys. Rev. Lett.* **68**, 1811.
- Wheeler, J.A., 1968, Superspace and the Nature of Quantum Geometrodynamics, in *Batelle Rencontres*, edited by C.M. DeWitt and J.A. Wheeler. Benjamin, New York.
- Zardecki, A., 1983, Modelling in Chaotic Relativity, *Phys. Rev. D* **28**, 1235.
- Zel'dovich, Ya.B. and Novikov, I.D., 1983, *Relativistic Astrophysics Vol.2, The structure and Evolution of the Universe*. The University of Chicago Press, in particular §22.

INTEGRABILITY OF THE MIXMASTER UNIVERSE

G. Contopoulos†, B. Grammaticos‡, and A. Ramani§

†Dept. of Astronomy, University of Florida
Gainesville, FL 32611, USA
and Dept. of Astronomy, University of Athens
Panepistimiopolis, GR 15783 Athens, Greece

‡LPN, Université Paris VII, Tour 24, 5^eme étage
75251 Paris, France

§CPT, Ecole Polytechnique, CNRS UPR 14,
91128 Palaiseau, France

Abstract. We show that the Mixmaster Universe, or Bianchi IX, model passes the Painlevé test, i.e. the solutions of the equations of motion have only poles as movable singularities. Thus this system is probably integrable and therefore non-chaotic.

1. INTRODUCTION

The Mixmaster Universe model, or Bianchi IX model was introduced by Belinski and Khalatnikov (1969) and independently by Misner (1969). This model is homogeneous but anisotropic: it expands along two directions and contracts along the third one. As one approaches the initial singularity, backwards in time, the directions of expansion and contraction change an infinite number of times. This system was believed to be ergodic and mixing (Ryan and Shepley 1975, Misner et al. 1977) i.e. chaotic.

If the Mixmaster Universe is mixing then it might connect the various parts of the early Universe and avoid the “causality problem” without introducing inflation. This fact explains the renewed interest of this model in recent years. However it should be noted that most of the oscillations of the Mixmaster Universe should occur before the Planck time, in which case the applicability of the General Theory of Relativity is questionable. But at any rate the problem is of theoretical interest in the framework of Relativity Theory.

In order to check the chaotic character of the Mixmaster Universe, several people calculated the maximal Lyapunov characteristic number (LCN). The first calculations (Zardecki 1983, Francisco and Matsas 1988) found a positive maximal LCN, but more accurate calculations (Hobill et al. 1989, Burd et al. 1990, Berger 1991a, Hobill 1991, Hobill et al., 1992) have shown that probably the maximal LCN is zero. Thus it seems that the Mixmaster Universe is not chaotic. Recently a number of efforts have been made to redefine the LCN's or find other indicators of chaos, in order to save the chaotic character of the Mixmaster Universe. However these attempts have not produced any useful results.

In order to check the chaotic character of the Mixmaster Universe model we have used recently (Contopoulos et al., 1993) a completely different method, namely, we have checked whether this model has the Painlevé property, i.e. whether its solutions do not have critical singularities but only poles. Then the solutions are single-valued, and this is the main check of integrability. In order to perform the analysis we write the model in Hamiltonian form, apply the Ablowitz, Ramani and Segur (1980) algorithm and show that the Mixmaster Universe passes the Painlevé test. This fact is a strong indication (Bountis et al., 1982, van Moerbeke 1988, Ramani et al., 1989) that this model is integrable.

2. THE EQUATIONS OF MOTION

The solution of Einstein's equations in the case of the Mixmaster Universe, model are written:

$$\begin{aligned} 2\ddot{\alpha} &= \left(e^{2\beta} - e^{2\gamma}\right)^2 - e^{4\alpha} \\ 2\ddot{\beta} &= \left(e^{2\gamma} - e^{2\alpha}\right)^2 - e^{4\beta} \\ 2\ddot{\gamma} &= \left(e^{2\alpha} - e^{2\beta}\right)^2 - e^{4\gamma}. \end{aligned} \quad (1)$$

These equations can be derived from the Hamiltonian

$$H \equiv \frac{1}{16} \left(p_\alpha^2 + p_\beta^2 + p_\gamma^2 - 2p_\alpha p_\beta - 2p_\beta p_\gamma - 2p_\gamma p_\alpha \right) - e^{4\alpha} - e^{4\beta} - e^{4\gamma} + 2e^{2(\alpha+\beta)} + 2e^{2(\beta+\gamma)} + 2e^{2(\gamma+\alpha)} = 0 \quad (2)$$

where $p_\alpha = -4(\dot{\beta} + \dot{\gamma})$, $p_\beta = -4(\dot{\gamma} + \dot{\alpha})$, $p_\gamma = -4(\dot{\alpha} + \dot{\beta})$. Dots mean derivatives with respect to the logarithmic time τ which was introduced by Belinski and Khalatnikov (1969), and is related to the coordinate time t by the relation:

$$\tau = -\ln t. \quad (3)$$

The numeral value of the Hamiltonian H in the Mixmaster model is zero.

If we make the canonical change of variables

$$\begin{aligned} x &= 2\alpha, \quad y = 2\beta, \quad z = 2\gamma \\ p_x &= -(\dot{y} + \dot{z}), \quad p_y = -(\dot{z} + \dot{x}), \quad p_z = -(\dot{x} + \dot{y}) \end{aligned} \quad (4)$$

we find

$$\begin{aligned} 2\dot{x} &= p_x - p_y - p_z \\ 2\dot{y} &= p_y - p_z - p_x \\ 2\dot{z} &= p_z - p_x - p_y \end{aligned} \quad (5)$$

$$\begin{aligned} \dot{p}_x &= 2e^x(e^y + e^z - e^x) \\ \dot{p}_y &= 2e^y(e^z + e^x - e^y) \\ \dot{p}_z &= 2e^z(e^x + e^y - e^z) \end{aligned} \quad (6)$$

with

$$H \equiv \frac{1}{4} \left(p_x^2 + p_y^2 + p_z^2 - 2p_y p_z - 2p_z p_x \right) + e^{2x} + e^{2y} + e^{2z} - 2e^{x+y} - 2e^{y+z} - 2e^{z+x} = 0 \quad (7)$$

The variables x, y, z, p_x, p_y, p_z are canonical variables in the Hamiltonian H and the energy has the particular value zero.

3. THE PAINLEVÉ TEST

We introduce now new, noncanonical, variables

$$X = e^x, Y = e^y, Z = e^z \quad (8)$$

and rewrite the equations of motion as.

$$\begin{aligned} 2\dot{X} &= X(p_x - p_y - p_z) \\ 2\dot{Y} &= Y(p_y - p_z - p_x) \\ 2\dot{Z} &= Z(p_z - p_x - p_y) \end{aligned} \quad (9)$$

$$\begin{aligned} \dot{p}_x &= 2X(Y + Z - X) \\ \dot{p}_y &= 2Y(Z + X - Y) \\ \dot{p}_z &= 2Z(X + Y - Z) \end{aligned} \quad (10)$$

In these variables the second members are quadratic functions. We can now apply the Albowitz, Ramani and Segur (1980) algorithm to eqs.(9) and (10). This algorithm consists in three steps:

- a) We obtain the leading singular behaviours. This means that we determine the solutions around a pole $\tau = \tau_o$, in the form s^{m_i} , where $s = \tau - \tau_o$ and m_i are the most negative powers of the solutions. Then we find also the coefficients of the leading terms.
- b) We find the resonances. That means that we expand the solutions in Laurent series starting with the leading terms, and we obtain recurrence relations for the successive coefficients of the terms of powers s^{m_i+r} the resonances are the values of r for which the coefficients are not uniquely determined.

- c) We check the compatibility of the solution. In each resonance case we have to solve an inhomogeneous linear system. The compatibility condition means that this system can be solved despite the fact that the determinant of the unknowns is zero.

The leading terms are found by setting:

$$X = x_1 s^{m_1}, Y = x_2 s^{m_2}, Z = x_3 s^{m_3}, p_x = p_1 s^{n_1}, p_y = p_2 s^{n_2}, p_z = p_3 s^{n_3} \quad (11)$$

in eqs.(9) and (10), and equating the terms with the most negative power in s , in both the left, and right hand sides of these equations. These most negative powers are among the following:

$$\begin{aligned} m_1 x_1 s^{m_1-1} &= \frac{x_1}{2} s^{m_1} (p_1 s^{n_1} - p_2 s^{n_2} - p_3 s^{n_3}) \\ m_2 x_2 s^{m_2-1} &= \frac{x_2}{2} s^{m_2} (p_2 s^{n_2} - p_3 s^{n_3} - p_1 s^{n_1}) \\ m_3 x_3 s^{m_3-1} &= \frac{x_3}{2} s^{m_3} (p_3 s^{n_3} - p_1 s^{n_1} - p_2 s^{n_2}) \end{aligned} \quad (12)$$

$$\begin{aligned} n_1 p_1 s^{n_1-1} &= 2x_1 s^{m_1} (x_2 s^{m_2} + x_3 s^{m_3} - x_1 s^{m_1}) \\ n_2 p_2 s^{n_2-1} &= 2x_2 s^{m_2} (x_3 s^{m_3} + x_1 s^{m_1} - x_2 s^{m_2}) \\ n_3 p_3 s^{n_3-1} &= 2x_3 s^{m_3} (x_1 s^{m_1} + x_2 s^{m_2} - x_3 s^{m_3}) \end{aligned} \quad (13)$$

Without loss of generality we can assume $n_1 \leq n_2 \leq n_3$. We distinguish two main cases:

Case I: $n_1 < n_2 \leq n_3$

Subcase Ia:

The lowest possible exponents in eqs.(12) and (13) give:

$$n_1 = -1, n_2 = 0, n_3 = 0 \quad (14)$$

and

$$m_1 = -1, m_2 = 1, m_3 = 1. \quad (15)$$

As we shall see, this is the generic singular behaviour with 6 free parameters. The coefficients of the lowest order terms in eqs. (12) and (13) give

$$p_1 = -2, x_1 = \pm i \quad (16)$$

while x_2, x_3, p_2, p_3 remain free.

Then we look for the resonances by setting:

$$\begin{aligned} X &= x_1 s^{m_1} + \gamma_1 s^{m_1+r}, Y = x_2 s^{m_2} + \gamma_2 s^{m_2+r}, Z = x_3 s^{m_3} + \gamma_3 s^{m_3+r} \\ p_x &= p_1 s^{n_1} + \delta_1 s^{n_1+r}, p_y = p_2 s^{n_2} + \delta_2 s^{n_2+r}, p_z = p_3 s^{n_3} + \delta_3 s^{n_3+r} \end{aligned} \quad (17)$$

in eqs.(9) and (10). The coefficients of the terms linear in (γ_i, δ_i) give:

$$\begin{aligned} \gamma_1(r-1) &= -\gamma_1 + \frac{x_1}{2} \delta_1 \\ \gamma_2(r+1) &= \gamma_2 - \frac{x_2}{2} \delta_1 \\ \gamma_3(r+1) &= \gamma_3 - \frac{x_3}{2} \delta_1 \end{aligned} \quad (18)$$

$$\begin{aligned}
\delta_1(r-1) &= -4x_1\gamma_1 \\
\delta_2 r &= 0 \\
\delta_3 r &= 0
\end{aligned} \tag{19}$$

The determinant of the coefficients of (γ_i, δ_i) for the resonances is zero. Thus

$$\left[\begin{array}{cccccc} r & 0 & 0 & -\frac{x_1}{2} & 0 & 0 \\ 0 & r & 0 & \frac{x_2}{2} & 0 & 0 \\ 0 & 0 & r & \frac{x_3}{2} & 0 & 0 \\ 4x_1 & 0 & 0 & r-1 & 0 & 0 \\ 0 & 0 & 0 & 0 & r & 0 \\ 0 & 0 & 0 & 0 & 0 & r \end{array} \right] = (r+1)r^4(r-2) = 0 \tag{20}$$

It is well known that the resonance -1 is related to the freedom of the location r_0 of the singularity. The quadruple 0 resonance is related to the free x_2, x_3, p_2, p_3 parameters below. It remains thus to investigate whether the $r = 2$ resonance satisfies the compatibility condition by expanding all variables up to terms linear in s and substituting in the full system (9) and (10).

Thus we set

$$\begin{aligned}
X &= \pm \frac{i}{s} + x_0 + \gamma_1 t + \dots \\
Y &= x_2 s + \dots \\
Z &= x_3 s + \dots \\
p_x &= -\frac{2}{s} + p_{10} + p_{11} t + \dots \\
p_y &= p_2 + p_{21} t + \dots \\
p_z &= p_3 + p_{31} t + \dots
\end{aligned} \tag{21}$$

in eqs. (9) and (10), and equate the coefficients of all the terms up to first order in s . Then we find

$$\begin{aligned}
x_0 &= 0, \quad p_{11} = \pm 2i(x_2 + x_3 - 2\gamma_1) \\
p_{10} &= p_2 + p_3, \quad p_{21} = \pm 2ix_2, \quad p_{31} = \pm 2ix_3.
\end{aligned} \tag{22}$$

There is one more equation beyond the number of undetermined coefficients, and this happens to be satisfied with the values (22). The fact that this equation is satisfied is the compatibility condition. If this equation were not satisfied then we would be forced to add logarithmic terms in the expansions (21), and this would mean that the original system would not be integrable.

Thus we have the following singular expansions:

$$\begin{aligned}
X &= \pm \frac{i}{s} + 0 + \gamma_1 s + \dots \\
Y &= x_2 s + \dots \\
Z &= x_3 s + \dots \\
p_x &= -\frac{2}{s} + (p_2 + p_3) \pm 2i(x_2 + x_3 - 2\gamma_1) s + \dots \\
p_y &= p_2 \pm 2ix_2 s + \dots \\
p_z &= p_3 \pm 2ix_3 s + \dots
\end{aligned} \tag{23}$$

The free parameter γ_1 enters at $r = 2$ and the Laurent series can be extended step by step, without any further problem. This expansion is generic, i.e. it has 6 free parameters. Thus the solution is of Painlevé type i.e. it has only poles as singularities. There are two further subcases of Case I.

Subcase Ib:

$$n_1 = -1, \quad n_2 = 0, \quad n_3 = 1. \quad (24)$$

This case corresponds to taking the free quantity p_3 equal to zero in Subcase Ia. Then

$$m_1 = -1, \quad m_2 = m_3 = 1, \quad n_3 = 1 \quad (25)$$

The resonances are the roots the equation

$$r^3(r+1)^2(r-2) = 0 \quad (26)$$

and the singular solution is

$$\begin{aligned} X &= \pm \frac{i}{s} + 0 + \gamma_1 s + \dots \\ Y &= x_2 s + \dots \\ Z &= x_3 s + \dots \\ p_x &= -\frac{2}{s} + p_2 \pm 2i(x_2 + x_3 - 2\gamma_1)t + \dots \\ p_y &= p_2 \pm 2ix_2 s + \dots \\ p_z &= \pm 2ix_3 s + \dots \end{aligned} \quad (27)$$

Subcase Ic:

$$n_1 = -1, \quad n_2 = n_3 = 1. \quad (28)$$

corresponding to both p_2 and p_3 vanishing in Subcase Ia. Then

$$m_1 = -1, \quad m_2 = m_3 = n_2 = n_3 = 1, \quad (29)$$

The resonances are the roots of the equation

$$r^2(r+1)^3(r-2) = 0 \quad (30)$$

and the singular solution is a special case of (23) with $p_2 = p_3 = 0$.

Next we examine cases with $n_1 = n_2$. We can first prove that there is no singular solution with $n_1 = n_2 < n_3$. Thus we check the

Case II: $n_1 = n_2 = n_3$

The same analysis as in the generic case leads to

$$\begin{aligned} n_1 &= n_2 = n_3 = -1 \\ m_1 &= m_2 = m_3 = -1 \\ x_1 &= x_2 = x_3 = \pm i \\ p_1 &= p_2 = p_3 = 2 \end{aligned} \quad (31)$$

The resonances are straight forward to compute and we find:

$$(r+1)^3(r-2)^3 = 0 \quad (32)$$

The (-1) resonance is again associated with the freedom of τ_0 . No logarithmic terms enter at this resonance.

On the other hand we must check whether the triple 2 leads to incompatibilities. Expanding all x_i, p_i 's up to first order in s we have checked that compatibility at $r = 2$ is indeed satisfied and we get

$$\begin{aligned} X &= \pm \frac{i}{s} + \gamma_1 s + \dots \\ Y &= \pm \frac{i}{s} + \gamma_2 s + \dots \\ Z &= \pm \frac{i}{s} + \gamma_3 s + \dots \\ p_x &= \frac{2}{s} \pm 2i(\gamma_2 + \gamma_3)s + \dots \\ p_y &= \frac{2}{s} \pm 2i(\gamma_3 + \gamma_1)s + \dots \\ p_z &= \frac{2}{s} \pm 2i(\gamma_1 + \gamma_2)s + \dots \end{aligned} \quad (33)$$

This solution is also of Painlevé type i.e. it has only poles as singularities. However the solution (33) depends on 4 free parameters only, i.e. it is nongeneric. At any rate the Mixmaster Universe model satisfies the Painlevé criterion of integrability, for all its singular solutions, generic and non-generic. We conclude that the Mixmaster model passes the Painlevé test. This is true whether the Hamiltonian (7) has the value zero or not.

4. ZIGLIN'S THEOREM

Ziglin's (1983) theorem is used to find whether a system is nonintegrable. Simplified expositions of this theorem were given by Yoshida et al. (1987) and Yoshida (1988). In order to apply this theorem we must find a linear periodic solution of the system. Then we find the Normal Variational Equations for this solution, i.e. the variational equations for deviations perpendicular to the linear solution. Thus we derive the monodromy matrices of the linearized variational equations, corresponding to closed loops around the periodic solution in the complex-time plane. The essence of Ziglin's theorem, is that if some solutions of the variational equations after such loops in the complex-time plane give different values then the system cannot be integrable. However if the monodromy matrices satisfy certain simple conditions then there is a *possibility* (not a proof) that the system is integrable.

In the present case we bring the Hamiltonian (7) to a standard form, by using a complex canonical change of variables

$$\begin{aligned} x &= iu + 2v \\ y &= iu - v + \sqrt{3}w \\ z &= iu - v - \sqrt{3}w \\ p_u &= i(p_x + p_y + p_z) \\ p_v &= 2p_x - p_y - p_z \\ p_w &= \sqrt{3}(p_y - p_z) \end{aligned} \quad (34)$$

Then the Hamiltonian (7) becomes

$$H = \frac{1}{12} (p_u^2 + p_v^2 + p_w^2) + e^{2iu} [e^{4v} + e^{-2(v-\sqrt{3}w)} + e^{-2(v+\sqrt{3}w)} - 2e^{-2v} - 2e^{v-\sqrt{3}w} - 2e^{v+\sqrt{3}w}] \quad (35)$$

This is of “Toda-type” (but not the Toda Hamiltonian itself). The corresponding equation of motion are

$$\begin{aligned} \ddot{u} &= -\frac{i}{3} e^{2iu} [e^{4v} + e^{-2(v-\sqrt{3}w)} + e^{-2(v+\sqrt{3}w)} - 2e^{-2v} - 2e^{v-\sqrt{3}w} - 2e^{v+\sqrt{3}w}] \\ \ddot{v} &= -\frac{1}{3} e^{2iu} [2e^{4v} - e^{-2(v-\sqrt{3}w)} - e^{-2(v+\sqrt{3}w)} + 2e^{-2v} - e^{v-\sqrt{3}w} - e^{v+\sqrt{3}w}] \\ \ddot{w} &= -\frac{1}{\sqrt{3}} e^{2iu} [e^{-2(v-\sqrt{3}w)} - e^{-2(v+\sqrt{3}w)} + e^{v-\sqrt{3}w} - e^{v+\sqrt{3}w}] \end{aligned} \quad (36)$$

These equations possess an obvious straight-line solution:

$$v = w = p_v = p_w = 0. \quad (37)$$

This leads to the equation

$$\ddot{u} = ie^{2iu}, \quad (38)$$

which can be integrated once to give

$$\dot{u}^2 - e^{2iu} = C \quad (39)$$

If we put now

$$U = i\sqrt{C}e^{iu}, \quad (40)$$

we find

$$U = \cos [\sqrt{C}(\tau - \tau_0)] \quad (41)$$

This is a simply periodic solution with period $2\pi/\sqrt{C}$ and has two simple zeros per period, which correspond to double poles in the potential.

A Ziglin’s problem with this precise structure has been studied by Yoshida et al. (1987). (This paper gives a generalization of Ziglin’s theorem to cases where the monodromy matrices are not diagonalizable).

The variational equations are the linearized equations (36) with u, v , and w replaced by $u + \xi, 0 + \eta$ and $0 + \zeta$, where u is the periodic solution derived from eqs. (40) and (41). We have

$$\begin{aligned} \ddot{\xi} &= -2e^{2iu}\xi \quad (\text{tangential}) \\ \ddot{\eta} &= -2e^{2iu}\eta \\ \ddot{\zeta} &= -2e^{2iu}\zeta \quad \} \quad (\text{normal}) \end{aligned} \quad (42)$$

Thus the normal variational equations in this case are identical to the tangential one. The tangential variational equations can be derived directly by linearization of eq. (39), which is a one-degree of freedom Hamiltonian. Therefore the normal variational equations have the same monodromy matrix, and cannot lead to non-single-valued solutions.

This remark is sufficient to prove that Ziglin’s theorem cannot prove nonintegrability of the Mixmaster Universe. But even if we can apply Ziglin’s theorem, we cannot prove integrability. We have only one more *indication* that the Mixmaster model is integrable.

5. CONCLUSIONS

The Mixmaster Universe was proposed as a chaotic model for the early Universe. However a detailed computation of the maximal Lyapunov Characteristic Number (LCN) did not lead to positive values, as one would expect in the case of chaos, if one uses the τ -time of Belinski and Khalatnikov (1969) that gives $\tau = \infty$ at the Big Bang.

Thus certain authors (Ferraz and Francisco 1991, Ferraz et al. 1991, Berger 1991b, Pullin 1991) introduced new time variables in order to derive a positive LCN. However such a method is ambiguous. In fact, even an integrable system, which is known to be nonchaotic, can give a positive LCN by an appropriate time transformation.

Consider a circular Kepler motion in which the azimuth θ is proportional to the time, $\theta = \Omega t$. A variation of Ω by $\delta\Omega$ gives a deviation

$$\eta = \delta\Omega t. \quad (43)$$

Thus an initial deviation $\eta_0 = \delta\Omega t_0$ grows linearly in time t . The corresponding LCN is zero

$$LCN = \lim_{t \rightarrow \infty} \frac{\ln(\frac{\eta}{\eta_0})}{t} = 0 \quad (44)$$

But if we replace this time by another time t' through the relation $t = e^{qt'}$ with $q > 0$ we find a positive Lyapunov characteristic number

$$LCN' = q > 0 \quad (45)$$

But in the new time t' the length of a circular arc, described in a time interval $\Delta t'$, increases as t' increases. Thus we have a phenomenon very similar to Zeno's paradox (Achilles and tortoise).

The question is, what is a physically appropriate time? In the circular Keplerian case it is obvious that the length of the circular arc itself is the best "time" to be used. In a similar way in a conservative Hamiltonian system expressed in action-angle variables the angles give an appropriate measure of time. (In the integrable cases they are linear functions of time).

In the case of the Mixmaster Universe the t -time is not appropriate for defining a LCN, because it terminates at $t = 0$. On the other hand the τ -time of Belinski and Khalatnikov (1969) tends to $+\infty$ when $t \rightarrow 0$, and in this time the equations of motion are derived from a conservative Hamiltonian. Thus the τ -time may be appropriate for checking whether the Mixmaster Universe is integrable or chaotic.

The fact that the Mixmaster model passes the Painlevé test in τ -time is a strong indication of integrability (Bountis et al. 1982, Ramani et al. 1989). In fact the solutions of the equations of motion are single-valued around every singularity, and this indicates that the orbits have no chaos.

The distinction between single-valued (or multiply-valued) and infinitely-valued integrals of motion is basic in Celestial Mechanics. Single-valued integrals are called "isolating" (Wintner 1941) and provide constraints on the possible forms of the orbits, thus avoiding chaos.

Our Painlevé analysis indicates that there probably are 2 more integrals besides the Hamiltonian itself for any value of H . We have tried to find polynomial integrals of low degree, but without success. We conclude that the integrals are either higher order

polynomials or functions of a different form. Thus the problem of finding integrals of motion for the Mixmaster Universe model is an important open problem.

But even if the Mixmaster model is proven definitely to be integrable there should be models deviating from it, that are mostly chaotic. This conclusion is based on the known fact that nonintegrable dynamical systems are generic, while integrable systems are exceptional. Therefore there is good hope that the effort to find truly chaotic Hamiltonian systems for the Universe will be satisfied.

REFERENCES

- Ablowitz, M. J., Ramani, A. and Segur, H., 1980, *J. Math. Phys.*, **21**, 715.
- Belinski, V. A. and Khalatnikov, I. M., 1969, *Sov. Phys. JETP*, **29**, 911 and **30**, 1174.
- Berger, B. K., 1991a, *Class. Quantum Grav.*, **7**, 203.
- Berger, B. K., 1991b, *Gen. Rel. Grav.*, **23**, 1385.
- Bountis, A., Segur, H. and Vivaldi, F., 1982, *Phys. Rev.*, **A25**, 1257.
- Burd, A. B., Buric, N. and Ellis, G. F. R., 1990, *Gen. Rel. Grav.*, **22**, 349.
- Contopoulos, G., Grammaticos, B. and Ramani, A., 1993, *J. Phys. A.*, (in press).
- Ferraz, K. and Francisco, G., 1991, *Phys. Rev.*, **D45**, 1158.
- Ferraz, K., Francisco, G. and Matsas, G. E. A., 1991, *Phys. Lett.*, **A156**, 407.
- Francisco, G. and Matsas, G.E. A., 1988, *Gen. Rel. Grav.*, **20**, 1047.
- Hobill, D., 1991, in *Nonlinear problems in Relativity and Cosmology*, S. L. Detweiler and J. R. Ipser, eds., N. Y. Acad. Sci. **631**, p. 15.
- Hobill, D., Bernstein, D., Simpkins, D. and Welge, M., 1989, in *Proc. 12th Int. Conf. Gen. Rel. Grav.*, Boulder, Univ. of Colorado, abstracts p.337.
- Hobill, D., Bernstein, D., Simpkins, D. and Welge, M., 1992, *Class. Quantum Grav.*, **8**, 1155.
- Landau, L. D. and Lifshitz, E. M., 1975, *The Classical Theory of Fields*, Pergamon Press, Oxford, eqs. 118.5-118.9.
- Misner, C. M., 1969, *Phys. Rev. Lett.*, **22**, 1071.
- Misner, C. M., Thorne, K. and Wheeler, J. A., 1977, *Gravitation*, Freeman, San Francisco.
- Pullin, J., 1991, in *SILARG VII Relativity and Gravitation, Classical and Quantum*, J. C. D'Olivio, E. Nahmad-Achar, M. Rosenbaum, M. P. Ryan, Jr., L. F. Urrutia and F. Zerrucke, eds., World Scientific, Singapore, p. 189.
- Ramani, A., Grammaticos, B. and Bountis, A., 1989, *Phys. Rep.*, **180**, 159.
- Ryan, M. P., Jr. and Shepley, L. C., 1975, *Homogeneous Relativistic Cosmologies*, Princeton Univ. Press, Princeton.
- van Moerbeke, P., 1988, in *Finite Dimensional Integrable Nonlinear Dynamical Systems*, P. G. L. Leach and W.H. Steeb, eds., World Scientific, Singapore, p.1.
- Yoshida, H., 1988, in *Finite Dimensional Integrable Nonlinear Dynamical Systems*, P. G. L. Leach and W.H. Steeb, eds., World Scientific, Singapore, p.74.
- Yoshida, H., Hietarinta, J., Grammaticos, B. and Ramani, A., 1987, *Physica*, **A144**, 310.
- Wintner, A., 1941, *The Analytical Foundations of Celestial Mechanics*, Princeton Univ. Press, Princeton.
- Zardecki, A., 1983, *Phys. Rev.*, **D28**, 1235.
- Ziglin, S. L., 1983, *Funct. Anal. Appl.*, **16**, 181 and **17**, 6.

CONTINUOUS TIME DYNAMICS AND ITERATIVE MAPS OF ELLIS-MACCALLUM-WAINWRIGHT VARIABLES

Teviet D. Creighton¹ and David W. Hobill²

Dept. of Physics and Astronomy, University of Calgary
Calgary, Alberta, T2N 1N4, Canada

Abstract. Solutions to Bianchi IX cosmologies using Ellis-MacCallum-Wainwright (“expansion normalized”) variables are examined. It is shown that an infinite number of periodic solutions exist. An iterative map (originally described by Bogoyavlenski) that describes the “late time” dynamics of the Bianchi IX model is also studied and its dynamics is compared to the continuous time case for the full Einstein equations. New variables are introduced that keep the phase space compact.

1. INTRODUCTION

The behavior of the Bianchi VIII and IX cosmologies near the cosmological singularity(ies) has attracted much attention recently, particularly in light of the controversy over whether or not the nonlinear evolution of the gravitational field as governed by Einstein’s equations can exhibit deterministic chaos.

Since the Bianchi models are spatially homogeneous, the Einstein field equations automatically reduce to a coupled set of ordinary differential equations (ODE’s) and researchers have naturally turned to the theory of nonlinear dynamical systems (which now refers to the study of the qualitative behavior of ODE’s and iterative maps) in order to attempt to understand the nature of time dependent gravity.

Historically a number of different approaches have been taken to study the behavior of the Bianchi cosmologies and the main ones are discussed in this volume. Many of the methods that have been employed are based upon approximating the full set of

¹email:tdcreigh@acs.ucalgary.ca

²email:hobill@scri.crag.ucalgary.ca

ODE's resulting from the Einstein equations with iterative maps that correspond to the dynamical behavior of the ODE's at late times when the transient behavior can be ignored. The question of whether these approximations are valid and if they are, what the range of their validity is, has been a subject of much study. In this contribution the approach of Ellis-MacCallum and Wainwright is taken in an attempt to shed light on the dynamics of the Mixmaster cosmology and the possibility that it may exhibit chaotic behavior.

2. EXPANSION NORMALIZED VARIABLES

Following upon the work of Ellis and MacCallum [1] and Wainwright and Hsu [2] we present a brief review of the basic variables and formalism used in this approach. Introducing a group invariant orthonormal tetrad, $\{\mathbf{e}_0 = \partial/\partial t, \mathbf{e}_i\}$ where Latin indices running from 1 to 3 represent spatial coordinates (Greek indices representing space-time coordinates run from 0 to 3), the basic variables are the non-zero commutators associated with the tetrad. For the diagonalized Bianchi models in Class A (see [3, 4]):

$$[\mathbf{e}_0, \mathbf{e}_i] = \theta_i(t) \mathbf{e}_i \quad (1)$$

where in eq. (1) there is no summation over the repeated indices. One also has commutation relations among the spatial tetrad components:

$$[\mathbf{e}_1, \mathbf{e}_2] = n_3(t) \mathbf{e}_3, \quad [\mathbf{e}_2, \mathbf{e}_3] = n_1(t) \mathbf{e}_1, \quad [\mathbf{e}_3, \mathbf{e}_1] = n_2(t) \mathbf{e}_2 \quad (2)$$

and the raising and lowering of spatial indices is accomplished with δ^{ik} . The $n_i(t)$ functions determine the spatial curvature (and therefore the Bianchi type of the isometry group).

The coefficients $\theta_i(t)$ are eigenvalues of the expansion tensor, θ_{jk} of the fluid source. The shear tensor σ_{ij} can be constructed from the expansion in the following manner:

$$\sigma_{ij} = \theta_{ij} - \frac{1}{3} \delta_{ij} \theta \quad (3)$$

where $\theta = \theta_i^i$ is the expansion scalar.

Since the shear is trace-free there are two independent diagonal components and it is more convenient to introduce the following shear variables:

$$\sigma_+ = \frac{3}{2}(\theta_2 + \theta_3 - \frac{2}{3}\theta), \quad \sigma_- = \frac{\sqrt{3}}{2}(\theta_2 - \theta_3) \quad (4)$$

The Einstein field equations can be written in the form of an autonomous system of ODE's with a six-dimensional phase space:

$$\frac{d\vec{x}}{dt} = \vec{F}(x)$$

where $\vec{x} = (\theta, \sigma_+, \sigma_-, n_i)$ are the phase space variables.

The system is invariant under scale transformations, ($\mathcal{X} = \lambda x$, $d\tau = \lambda dt$). Therefore one may introduce dimensionless variables by factoring out the overall expansion. Since \vec{x} typically diverges as one approaches a cosmological singularity at $t = 0$, this

process attempts to distinguish between different asymptotic states of the phase space variables near the “big bang”.

Defining the dimensionless (expansion-normalized) shear components:

$$\Sigma_{\pm} = \frac{\sigma_{\pm}}{\theta} \quad (5)$$

and the dimensionless (expansion-normalized) spatial curvature components:

$$N_i = \frac{n_i}{\theta} \quad (6)$$

along with a dimensionless time parameter τ , one proceeds along the lines used to study the homogeneous and isotropic Friedmann-Lemaître-Robertson-Walker cosmologies by defining a cosmological length scale-factor $R(t)$ by $\tau = \ln R(t)$. A Hubble parameter H is then defined as:

$$3H = 3\dot{R}/R = \theta \quad (7)$$

where the overdot represents d/dt . The relationship between t and τ is then:

$$\frac{dt}{d\tau} = \frac{3}{\theta}. \quad (8)$$

The deceleration parameter q is defined as

$$q = -\frac{R\ddot{R}}{\dot{R}^2} \quad (9)$$

while the density parameter Ω is

$$\Omega = \frac{3\rho}{\theta^2} \quad (10)$$

where ρ is the physical mass density.

The scale invariance reduces the six-dimensional phase space to one of five dimensions with variables $\vec{\mathcal{X}} = (\Sigma_+, \Sigma_-, N_i)$ and the equations remain as an autonomous set of ODE's in the form:

$$\frac{d\vec{\mathcal{X}}}{d\tau} = \vec{\mathcal{F}}(\mathcal{X}).$$

Explicitly one can write the equations as [2]:

$$\begin{aligned} N'_1 &= (q - 4\Sigma_+)N_1 \\ N'_2 &= (q + 2\Sigma_+ + 2\sqrt{3}\Sigma_-)N_2 \\ N'_3 &= (q + 2\Sigma_+ - 2\sqrt{3}\Sigma_-)N_3 \\ \Sigma'_+ &= -(2 - q)\Sigma_+ - 3S_+ \\ \Sigma'_- &= -(2 - q)\Sigma_- - 3S_- \end{aligned} \quad (11)$$

where the primes ('') represent the derivatives with respect to τ . The following notation is used to simplify the ODE's:

$$S_+ = \frac{1}{2}[(N_2 - N_3)^2 - N_1(2N_1 - N_2 - N_3)]$$

$$S_- = \frac{\sqrt{3}}{2}[(N_3 - N_2)(N_1 - N_2 - N_3)].$$

The deceleration parameter becomes

$$q = \frac{1}{2}(3\gamma - 2)(1 - K) + \frac{3}{2}(2 - \gamma)\Sigma$$

where

$$\Sigma = \Sigma_+^2 + \Sigma_-^2$$

and

$$K = \frac{3}{4}[N_1^2 + N_2^2 + N_3^2 - 2(N_1N_2 + N_2N_3 + N_3N_1)]$$

is an expansion-normalized spatial curvature scalar and the constant γ is obtained from a perfect fluid equation of state in the form:

$$p = (\gamma - 1)\rho.$$

Since no non-trivial analytic solutions to the system of ODE's given above currently exist, one needs to use numerical methods to investigate the explicit time-dependent behavior of the spacetimes described by them. This set of equations form the basis of the work reported in this article.

3. NUMERICAL SIMULATIONS

When trying to numerically follow the curvature variables N_i as they approach the initial singularity, one has the difficulty that at least two of them tend toward zero while another one grows. Very quickly accuracy is lost and at late times in the variable τ the numerical code eventually crashes. One way to avoid this problem is to change dynamic curvature variables to the logarithms of the curvature components Z_i defined by:

$$N_i = \exp(-Z_i) \quad (12)$$

and by making a change of independent variable $\tau \rightarrow -\tau$ the following set of ODE's are useful for numerical studies of the EMW formalism:

$$\begin{aligned} Z'_1 &= q - 4\Sigma_+ \\ Z'_2 &= q + 2\Sigma_+ + 2\sqrt{3}\Sigma_- \\ Z'_3 &= q + 2\Sigma_+ - 2\sqrt{3}\Sigma_- \\ \Sigma'_+ &= (2 - q)\Sigma_+ + 3S_+ \\ \Sigma'_- &= (2 - q)\Sigma_- + 3S_- \end{aligned} \quad (13)$$

and S_+ , S_- , q , Σ and K are now written in terms of the functions Z_i . These equations have the advantage that the growth of the curvature variables is "not too rapid" and the singularity is approached as $\tau \rightarrow \infty$ thereby introducing a singularity avoiding time slicing. The disadvantage is that phase space is still not compact and the dynamics slows appreciably as one approaches the singularity which requires an ever increasing amount of computational time.

Since the vacuum Kasner solution (Bianchi Type I) plays an important role in the asymptotic behavior, one must express the Kasner solution in terms of the variables (Σ_{\pm} , Z_i). The Kasner metric is diagonal and written in terms of three-power law components:

$$ds^2 = dt^2 - (t^{2p_1}dx^2 + t^{2p_2}dy^2 + t^{2p_3}dz^2),$$

where the p_i 's must obey the relationships:

$$p_1 + p_2 + p_3 = 1 \quad \text{and} \quad p_1^2 + p_2^2 + p_3^2 = 1.$$

The relationship between the p_i 's and the expansion normalized variables are [2]:

$$\begin{aligned} p_1 &= \frac{1}{3}(1 - 2\Sigma_+) \\ p_2 &= \frac{1}{3}(1 + \Sigma_+ + \sqrt{3}\Sigma_-) \\ p_3 &= \frac{1}{3}(1 + \Sigma_+ - \sqrt{3}\Sigma_-) \end{aligned} \quad (14)$$

which can be written as $\Sigma = \Sigma_+^2 + \Sigma_-^2 = 1$. We will refer to this last equation as the “Kasner condition” which clearly defines a unit circle in the $\Sigma_+\Sigma_-$ plane and that circle is called the “Kasner ring”. The exact Bianchi Type I spacetime has vanishing spatial curvature and therefore $N_1 = N_2 = N_3 = 0$. Such solutions are equilibrium solutions to eqns. (11).

Figure 1a shows a typical evolution of the logarithmic curvature components, Z_i 's as a function of the time parameter τ . As in the case with the logarithmic Belinskii-Khalatnikov-Lifshitz (BKL) variables (α, β, γ) [5] the evolution of the Z_i 's are piece-wise linear in the time parameter. From Figure 1b (which presents the time dependence of Σ over the same interval of τ) it is clear that the Kasner condition is met during the time that the values of Z_i are linear functions of τ . The solutions are not true Kasner solutions (since that would require all of the N_i 's vanish), but it can be seen in Figure 2 which shows an evolution over long time intervals τ that the values of the logarithms of the curvature components take on arbitrarily large negative values as the simulation approaches the singularity. The numerical solution becomes approximately Kasner with ever increasing accuracy as time progresses.

Following the terminology of BKL one can define each Kasner **epoch** as the time interval over which the Bianchi IX solution has constant values of the parameters p_i and $\Sigma = 1$. A transition from one Kasner epoch to another occurs when the value of Σ is different from unity. A Kasner **era** is the time period over which one of the $Z_i(\tau)$ curvature components monotonically decreases. In Figure 1b there are two distinct Kasner eras (determined when the value of Σ falls below 3/7 [8]).

The spectrum of Lyapunov exponents is also calculated for the dynamics and the results for the evolution shown in Figure 1 is shown in Figure 3. Four of the exponents tend toward zero while another (and the sum) remain negative indicating a lack of chaos in these variables. The Lyapunov exponents are calculated by assuming that two or more orbits (phase space trajectories obtained as a solution to the ODE's) initially separated by a distance $d(0) = \delta_0$ will after a duration of time, t be separated by $d(t) = \delta_0 2^{\kappa t}$ where κ is called the Lyapunov exponent. There are N Lyapunov exponents in an N dimensional phase space, one each for the N possible orthogonal directions along which the separation of trajectories can occur. If $\kappa > 0$ then the system is said to be chaotic.

In practice one calculates the the Lyapunov exponents in the following manner: suppose the system of interest can be written in the form $dx_i(t)/dt = F_i(\vec{x})$, where $i = 1, N$ and the number of components of the vector \vec{x} is the phase space dimension. Introduce a small perturbation δx_i to the vector x_i . The evolution of the perturbations is governed by the linearized equations: $d(\delta x_k)/dt = J_k^i \delta x_i$ where $J_k^i = \partial F_k / \partial x_i$ is the

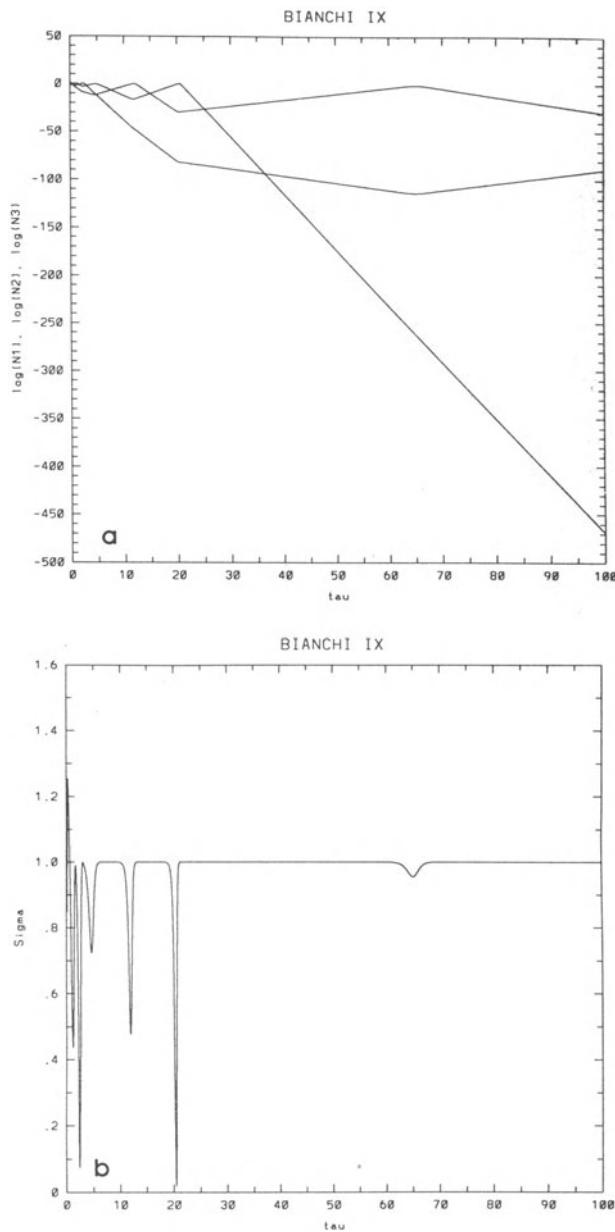


Figure 1. A time series plot of the evolution of the EMW variables. In (a) the time dependence of the logarithmic curvature variables is shown while (b) demonstrates that at late times the Kasner condition ($\Sigma = 1$) is met except when there is a transition from one Kasner epoch to another.

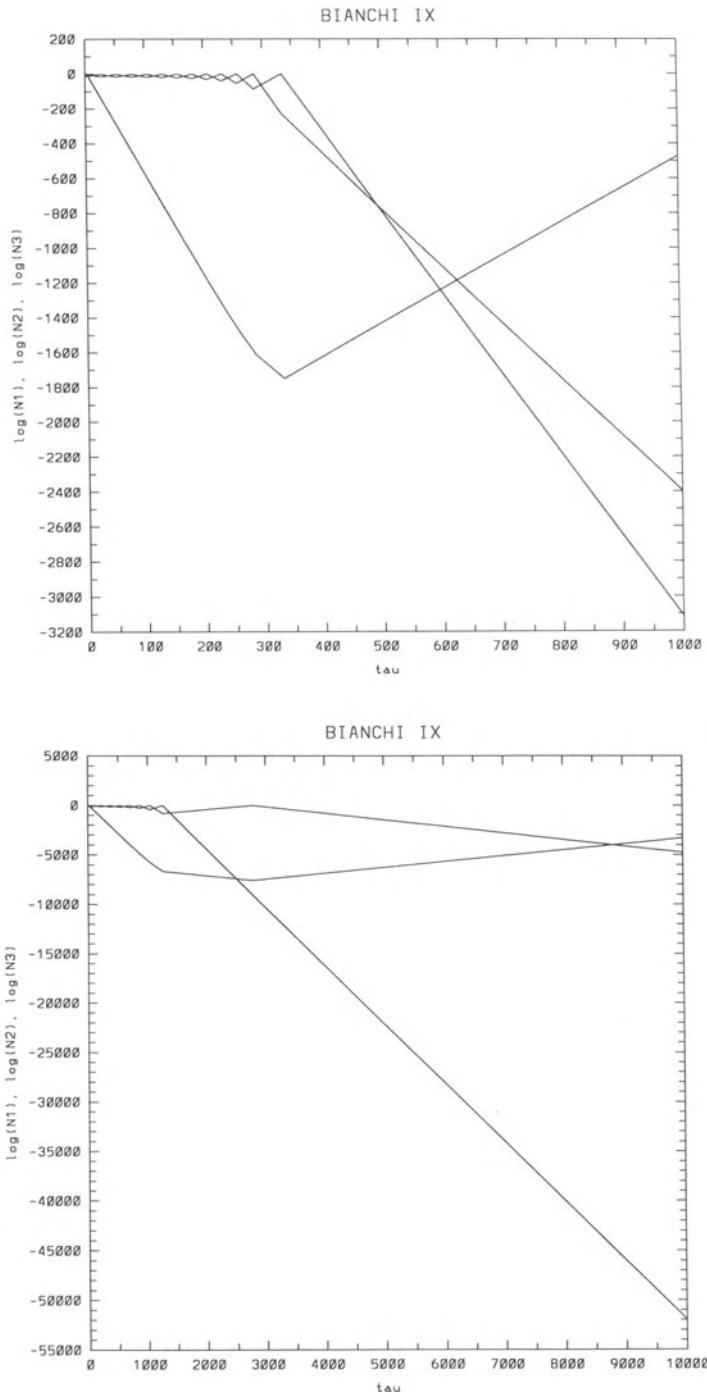


Figure 2. Typical evolutionary behavior of the curvature variables over much longer times than those shown in Figure 1 in the parameter τ . Notice that the $Z_i = \ln N_i$ take on large negative values indicating that $N_i \rightarrow 0$ on approach to the cosmological singularity ($\tau \rightarrow \infty$).

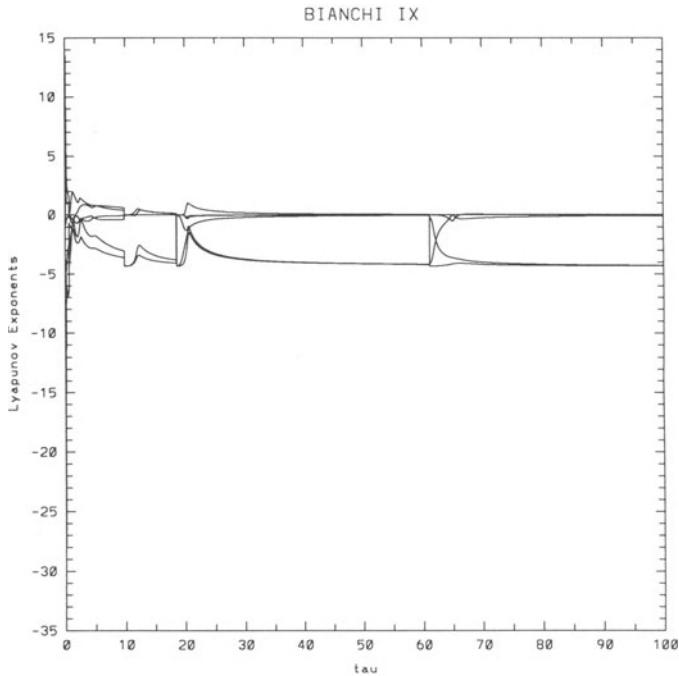


Figure 3. The Lyapunov exponents for the evolution shown in Figure 1. The entire spectrum plus the sum of the exponents is shown.

Jacobian matrix of the flow. In order to prevent the perturbation from growing without bound, the initial disturbance is ‘renormalized’ periodically and the perturbation equations are then recalculated. The Lyapunov exponents are defined in terms of the absolute values of the Jacobian matrix eigenvalues, j_i , averaged over the number n of renormalizations.

$$\kappa_i = \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 j_i(n).$$

In both the BKL and Hamiltonian descriptions of the Bianchi IX cosmology the transition from one Kasner era to another can be described by simpler iterative maps that do not depend upon a continuous time parameter. For the BKL formulation the map is algebraic and computes the parameter u associated with a particular Kasner solution in terms of the previous value of u [5]. In the Hamiltonian picture, the the bounce of an orbit off the expanding triangular potential [6] is determined by a transition rule in terms of the incident and reflected angles. Similarly the EMW formalism can be presented in the form of an iterative map. Since a Kasner solution is described by values of Σ_+ and Σ_- that sit on the Kasner ring, the transition from one Kasner era to another should be describable by the transition from one point on the Kasner ring to another. Bogoyavlenski [7] was able to find such a map and its equivalent in terms of the EMW variables has been described by Ma and Wainwright [8]. In the next section we study the dynamics of the so-called **B-map**.

4. THE DISCRETE KASNER MAPPING (B-MAP)

It is instructive to consider the evolution of the system (13) in the Σ_+, Σ_- phase subspace. Figure 4a shows a typical trajectory in this space as computed from the contin-

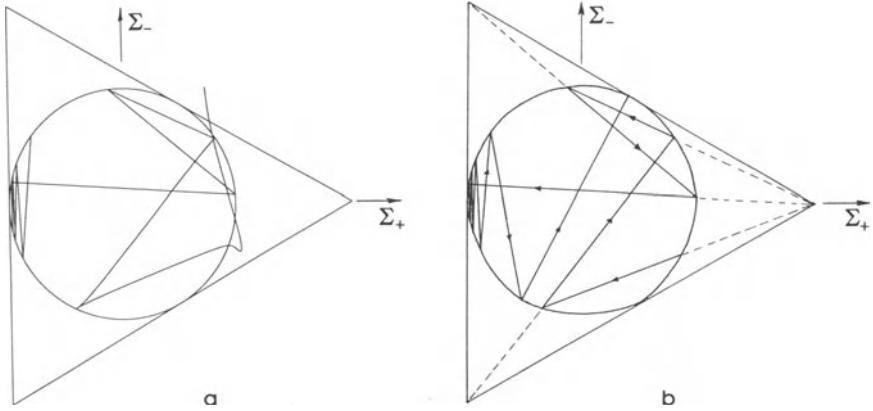


Figure 4. Phase space trajectory in the Σ_+ , Σ_- subspace, (a) computed numerically from the continuous ODE's [eqns. (13)] and (b) computed using the graphical construction of the B-map.

uous time dynamics. It can be seen that the initial curved trajectory gives way quite soon to a series of very nearly straight lines from point to point on the ring. At arbitrarily large timescales, when the transition time becomes negligible compared with the time spent on the Kasner ring (or arbitrarily close to it), these transition trajectories will become perfect secants.

Even more interesting is the fact that the trajectory during an “ideal” or instantaneous transition can be determined by a simple geometric construction. The construction involves three points $(2, 0)$, $(-1, \sqrt{3})$, $(-1, -\sqrt{3})$ on the Σ_+ , Σ_- plane, which can be viewed as the vertices of an equilateral triangle circumscribed about the Kasner ring. Given an initial point on the ring, the transition trajectory can be found by taking a line from the vertex nearest the initial point through that point and on until it reintersects the ring. The secant thus formed is the transition trajectory, and the reintersection point is thus the Kasner epoch following the transition. Also, the vertex used in this construction determines which of the variables Z_i triggered the transition: the vertices $(2, 0)$, $(-1, \sqrt{3})$, and $(-1, -\sqrt{3})$ correspond to Z_1 , Z_2 , and Z_3 respectively. When the next transition occurs, a similar construction can be performed on the new epoch to find the next Kasner epoch, and so on. Figure 4b shows an example of such a construction based upon initial values of Σ_+ and Σ_- determined from the continuous time dynamics. The B-map approximation to the full dynamics improves as time increases.

The transition rule given above also can be used to determine the change of an era. First one notices that the points where the equilateral triangle and the Kasner ring are tangent locate the equilibrium points of the B-map since those points are connected to two vertices by a straight line and therefore map into themselves. Any small deviation away from the equilibrium point will cause the map to begin its action and the transitions will oscillate near the equilibrium point until the positions of Σ_+ and Σ_- are a finite distance away from the equilibrium position. If one divides the Kasner circle into three equal triangles formed by drawing the bisectors of each vertex to their mutual intersection at the center of the Kasner ring, it will be seen that the oscillations of Kasner epochs close to an equilibrium point stay within one sub-triangle. Once one of the trajectories crosses the boundary of a sub-triangle, a new era is established and the oscillations of two different curvature components will continue until another sub-triangle boundary is crossed. (See Figure 5.)

In the large-timescale limit much of the information of a Kasner epoch is contained

in a single parameter, the angular position on the Kasner ring, whose value in any given epoch depends only on its value in the preceding epoch. We can therefore describe the evolution in a discrete, time-invariant manner using the B-map. The Z_i variables determine the time spent in each epoch, but not the sequence of epochs.

As mentioned earlier, a Kasner *era* consists of a succession of epochs in which a given pair of Z_i variables are oscillating. With the geometric construction above, this corresponds with transitions back and forth between a given pair of vertices of the circumscribed triangle. This in turn can be identified with particular $(2\pi/3)$ -radian arcs on the Kasner ring: epochs in $(0, 2\pi/3)$ involve oscillations between Z_1 and Z_2 , those in $(2\pi/3, 4\pi/3)$ involve Z_2 and Z_3 , and those in $(4\pi/3, 2\pi)$ involve Z_3 and Z_1 . Thus, in the large-timescale limit, an era consists of a succession of epochs lying in a given one of the above arcs; a transition of eras is a transition which crosses into a different arc.

Numerically, the B-map giving the transition from an angle θ_n to an angle θ_{n+1} is denoted by a function $\theta_{n+1} = f(\theta_n)$, where:

$$\theta_{n+1} = f(\theta_n) = \arccos \left(\frac{4 - 5 \cos \theta_n}{5 - 4 \cos \theta_n} \right), \quad 0 \leq \theta_n \leq \frac{\pi}{3}. \quad (15)$$

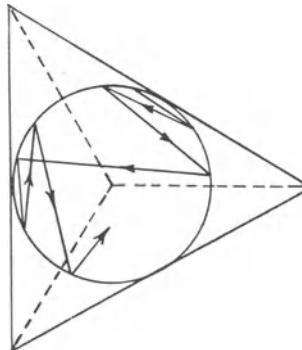


Figure 5. Determining transitions between Kasner eras (each containing a number of Kasner epochs) using the B-map.

For other values of θ_n , we compute θ_{n+1} by means of the symmetries of the map.

We note that each such wedge of angle $\pi/3$ maps into a wedge twice as large. This means that each point on the ring has two points which map to it via the B-map. Thus the B-map is not reversible, which is due to the fact that the angle on the ring contains only some of the information of the state. In general, the values of the Z_i variables are needed to project the system backwards in time. Figure 6 shows how the N_i 's change as the solution evolves by traversing the interior of the Kasner ring.

Furthermore, even when we apply the map in the normal (forward) direction, the fact that we are mapping smaller regions into larger ones means that we are losing information. If we know the angle θ to within a region of width $\Delta\theta$, and this region maps onto a larger region of size $k\Delta\theta$, $k > 1$, then we have lost $\log_2 k$ bits of precision in our knowledge of the predicted value of the angle. This loss of precision is the *Lyapunov exponent* of the transition, denoted by κ . For small $\Delta\theta$ we have $k = |f'(\theta)|$,

so the Lyapunov exponent as a function of initial angle is:

$$\kappa(\theta) = \log_2 \left(\frac{3}{5 - 4 \cos \theta} \right), \quad 0 \leq \theta \leq \frac{\pi}{3} \quad (16)$$

and symmetrically for other values of θ .

In general, the fixed points or “equilibria” of a map $f(\theta)$ are those points θ_0 such that $f(\theta_0) = \theta_0$. If a trajectory is initially a small distance $\Delta\theta_0$ from the equilibrium, then after n transitions it will be at a distance

$$\Delta\theta_n \approx \Delta\theta_0 2^{\kappa(\theta_0)}. \quad (17)$$

provided $\kappa(\theta)$ is approximately constant over $\Delta\theta$. So, if we treat n as a discretization of a continuous variable, we can write the rate of divergence (per transition) from equilibrium as:

$$\frac{\partial}{\partial n} \Delta\theta \approx \ln(2)\kappa(\theta_0)\Delta\theta. \quad (18)$$

More generally, the rate of divergence will be a power series in $\Delta\theta$, with no zeroth-order term.

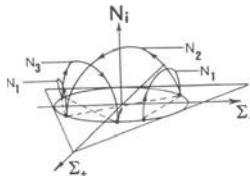


Figure 6. The action of the B-map as seen from the higher-dimensional phase space that includes the curvature variables, N_i .

If $\kappa(\theta_0) < 0$, the equilibrium point is stable; i.e. there is some finite range $\Delta\theta$ within which all trajectories converge on the equilibrium. If $\kappa(\theta_0) > 0$, the equilibrium point is unstable, and any trajectory not exactly on the equilibrium will eventually diverge from it. If $\kappa(\theta_0) = 0$, the equilibrium point is neutrally stable; the absolute convergence or divergence of nearby trajectories will depend on the nonlinear terms in the divergence expression above.

Geometrically, it can be seen that the B-map has fixed points at $\theta = (2n + 1)\pi/3$, i.e. where the circumscribed triangle is tangent to the Kasner ring. These equilibria are unstable, as can be seen using the geometric construction of the B-map: for any epoch a small distance from the equilibrium, each transition carries the system to epochs further and further from the equilibrium. However, the Lyapunov exponent is zero at these points, so trajectories close to an equilibrium will diverge at a rate proportional to the *square* of the distance from the equilibrium. Thus, while they are not stable in an absolute sense, they are neutrally stable as far as linear approximations go.

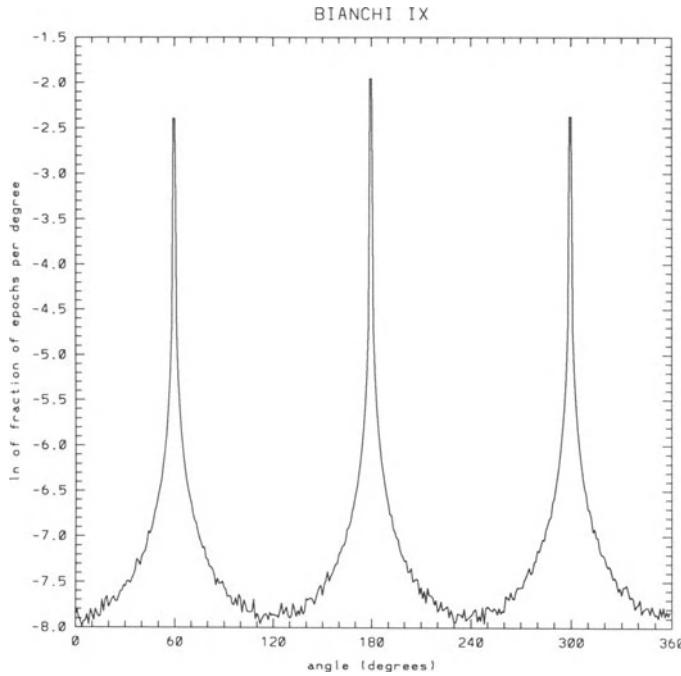


Figure 7. Population density of epochs per degree of arc on Kasner ring. Initial angle 0.8 ± 10^{-12} radians; 10^6 iterations.

For the B-map there is only a finite range in θ . We would expect a system to spend more epochs near to an equilibrium point than elsewhere, since a trajectory landing near an equilibrium point will be “trapped” there for several epochs before it moves away. This was demonstrated using a computer algorithm to iterate the B-map many times on some arbitrary initial epoch, and counting the number of times that the system enters epochs in various bins on the Kasner ring. The results for a typical run of 10^6 iterations and 1° bins is shown in Figure 7.

One feature of this distribution of epochs is that it is not trilaterally symmetric, as we would expect; i.e. the “time”-averaged distribution is not the same as the expected ensemble average (where “time” is referring to the number of transitions, not a physical time). From this we can conclude that even after 10^6 iterations the system has not become stationary; i.e. the distribution is still evolving in “time”. Indeed, a closer examination of the data revealed that for various initial angles, over 2×10^5 of the epochs occurred in a single era in which the system was stuck very near an equilibrium point. It is expected that, if we were to increase the number of iterations by a given factor, this would simply increase the probability that an era transition would move the system within some arbitrarily small neighbourhood of an equilibrium point, from which it would take an arbitrarily long time to escape, and would not bring us any closer to a stationary solution in the population density.

Another indication of this nonstationary behaviour is the evolution of the average Lyapunov exponent. Figure 8 shows the transition-averaged Lyapunov exponent computed after various numbers of transitions for a given trajectory. Although the Lyapunov exponent for any given transition of the Kasner map is positive, the average exponent appears to approach zero asymptotically, which could occur only if the system is spending a greater and greater number of epochs within arbitrarily small

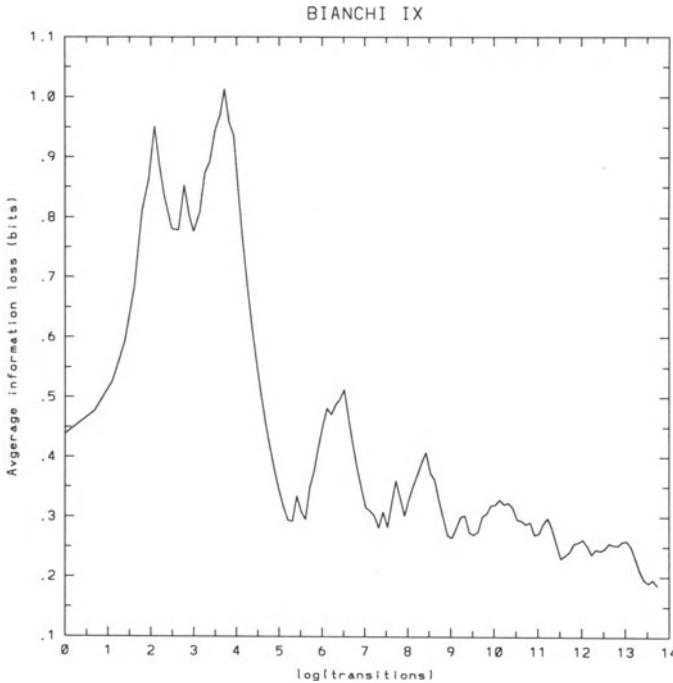


Figure 8. Cumulative transition-averaged Lyapunov exponent for a particular trajectory: Initial angle 0.8 ± 10^{-12} radians; 10^6 iterations in total.

neighbourhoods of the equilibrium points. If the ultimate average Lyapunov exponent for the system is indeed zero, then the final limit of the population density must be a set of Dirac δ -distributions centered on the equilibrium points. The ensemble average distribution would necessarily be three δ -distributions of equal height (since the ensemble must reflect the symmetries of the system), but the transition-averaged distribution might approach a single δ -distribution about one of the equilibria, which would depend on the initial data.

5. PERIODIC SOLUTIONS OF THE B-MAP AND COMPACTIFIED PHASE SPACE

From the geometric construction of the B-map it becomes obvious that not only do three equilibrium points exist but periodic solutions that cycle over $N > 2$ points exist. The symmetry of the problem does not allow non-degenerate $N = 2$ solutions since these correspond to the equilibrium points. Figure 9 shows the two $N = 3$ as well as one of the three $N = 4$ periodic B-map solutions. The period-3 solutions represent three era changes each one containing only one epoch while the period-4 solutions consist of two eras containing two epochs each. What do these periodic solutions correspond to in the full continuous time dynamics?

Figure 10 provides an answer for the period-3 B-map. Using the values of Σ_+ and Σ_- determined from the Kasner ring and appropriate values of the N_i 's, the evolution of the curvature variables is determined from the full ODE's. They clearly are not periodic functions of the time parameter τ . However there is a behavior that is close to being self-similar. Plotting the same evolution in terms of the logarithm of τ does give periodic behavior in the length of time for which the solution stays in a Kasner era.

Figure 11 shows the time-dependence of both the period-3 and the period-4 B-map solutions determined from the ODE's. These plots clearly show the periodicity in the

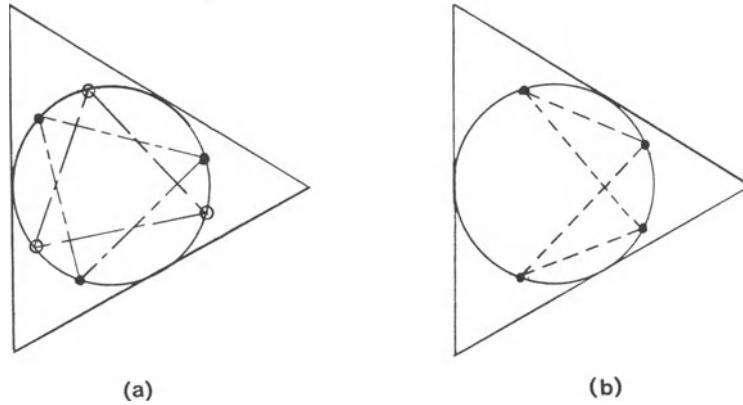


Figure 9. Periodic solutions of the B-map. Two period-3 solutions are shown in (a) while one period-4 solution is shown in B. The other period-4 solutions are obtained by a rotation through $2\pi/3$.

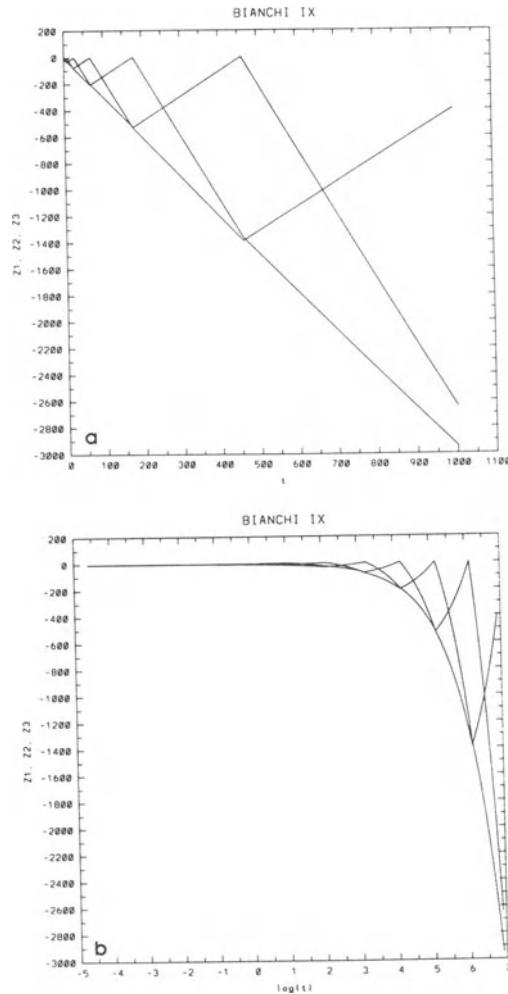


Figure 10. The continuous time dynamics of the period-3 B-map solution in terms of the parameter τ and in terms of $\ln \tau$.

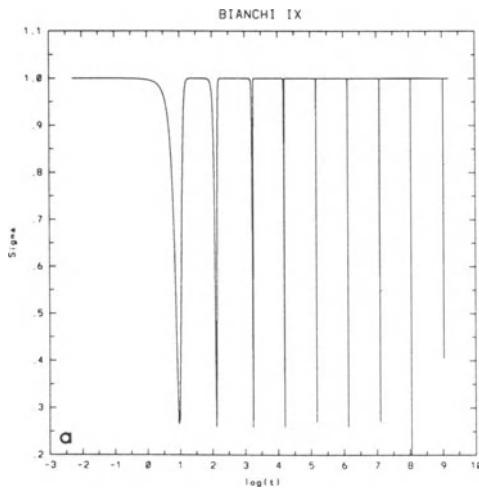


Figure 11. The continuous time dynamics of Σ for the period-3 (a) and the period-4 (b) B-maps as a function of $\ln \tau$.

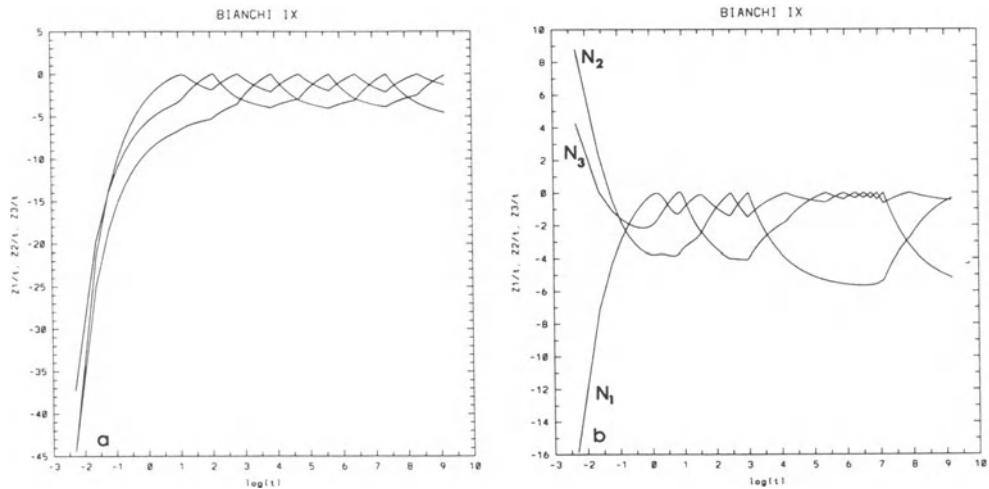


Figure 12. The compactified curvature variables (a) for a period-4 B-map evolution and (b) for a generic evolution that is non-periodic. The variables remain finite during the evolution but the singularity-avoiding time parameter still stretches the evolution near the singularity.

$\ln \tau$ parameterization. In addition it can be seen that each Kasner era in the period-4 solution has two Kasner epochs.

Finally it can be seen from the evolution of the curvature variables as a function of the parameter τ (Figure 10) that the increase in magnitude of those variables is bounded by a linear function of τ . This would suggest that phase space compactification could be accomplished by introducing variables N_i/τ . Figure 12 shows the compactification of these variables for the period-4 B-map solution as well as a generic case. In both cases, the values of the variables do stay finite. However since the time coordinate τ still stretches to infinitely large intervals as the singularity is approached, the Lyapunov exponents in these variables do not change from the case one had with the standard curvature variables.

ACKNOWLEDGEMENTS

The work reported here was financially supported by both a research grant and an Undergraduate Student Research Award from the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank Adrian Burd, Beverly Berger and particularly John Wainwright for stimulating discussions that enhanced our understanding of the dynamics of the Mixmaster cosmologies.

REFERENCES

- [1] Ellis G.F.R. and MacCallum M.A.H., (1969), *Comm. Math. Phys.* **12**, 108.
- [2] Wainwright, J., and Hsu, L.,(1989) *Class. Quan. Grav.*, **6**, 1409.
- [3] MacCallum, M., (1979), in *General Relativity (An Einstein Centenary Survey)*, (Hawking and Israel, eds), CUP, Cambridge, UK.
- [4] Jantzen, R., (1984) in *Cosmology of the Universe*, (Fang and Ruffini, eds), World Scientific, Singapore.
- [5] Belinskii, V.A., Khalatnikov, I.M. and Lifshitz, E.M., (1970), *Adv. Phys.*,**19**, 525.
- [6] Ryan, M.P. and Shepley, L.C., (1975), *Homogeneous Relativistic Cosmologies*, Princeton Univ. Press, Princeton, NJ.
- [7] Bogoyavlenski, O.I., (1985), *Methods in the Qualitative Theory of Dynamical Systems in Astrophysics and Gas Dynamics*, Springer Verlag, New York, NY.
- [8] Wainwright, J. and Ma, P., (1992) in *Relativity Today*, (Z. Perj's, ed), World Scientific, Teaneck, NJ.

A DYNAMICAL SYSTEMS APPROACH TO THE OSCILLATORY SINGULARITY IN BIANCHI COSMOLOGIES¹

P. K-H. Ma and J. Wainwright

Department of Applied Mathematics, University of Waterloo
Waterloo, Ontario, N2L 3G1 Canada

Abstract. We describe the behaviour of the orthogonal Bianchi cosmologies of types VIII and IX near the big-bang in terms of an attractor of a dynamical system. Comparisons are made with previous work.

1. INTRODUCTION

The behaviour of Bianchi cosmologies of types VIII and IX near the big-bang has attracted the attention of relativists since 1969, when it was revealed that the evolution into the past had an oscillatory nature (Belinskii and Khalatnikov (1969), Misner (1969)), now recognized as stochastic or chaotic (Khalatnikov et al (1985), Barrow (1982)). Nevertheless, this behaviour has not been fully understood from a mathematical point of view. Since the Einstein field equations (EFEs) for the Bianchi models can be written as an autonomous system of DEs, it is natural to use the qualitative theory of DEs, now referred to as the theory of dynamical systems (e.g. Nemytskii and Stepanov (1960), Anosov and Arnold (1988)) to investigate their evolution. In this paper, we use expansion-normalized variables to describe the behaviour of Bianchi cosmologies of types VIII and IX near the big-bang in terms of an attractor of a dynamical system. This approach provides a mathematically precise description of the behaviour, while at the same time having a simple physical and geometrical interpretation. The necessary background is contained in Wainwright and Hsu (1989) (subsequently referred to as WH) which should be read concurrently with this work.

¹This paper is a corrected version of the article appearing in *Relativity Today*, edited by Z. Perjés, Nova Science Publishers, Commack, NY, 1992

2. HISTORY

Three completely different approaches have been used to describe the asymptotic behaviour of the Bianchi models near the big-bang, which we refer to as follows:

1. The piecewise approximation approach
2. The Hamiltonian approach
3. The dynamical systems approach

In the first approach, the asymptotic behaviour is described in terms of the oscillations of the three independent length scales in the metric. This approach was pioneered by Belinskii and Khalatnikov (1969), and discussed in more detail by Belinskii, Khalatnikov and Lifshitz (1970), henceforth abbreviated as BKL.

BKL assert (pp. 532, 538) that in a model of Bianchi type VIII or IX, the matter density becomes dynamically negligible as $t \rightarrow 0$ and hence one can use the vacuum Einstein equations to analyze the early stages of these model universes. BKL then give heuristic arguments (p. 535) based on the form of the field equations to suggest that the evolution of a generic model into the past consists of an infinite sequence of time intervals called *epochs*, that can be grouped into different *eras*. The epochs and eras are defined by the behaviour of the three independent length scales in the metric. During an epoch, as $t \rightarrow 0$, two of the length scales decrease while the third increases in such a way that the solution can be approximated by a Kasner vacuum solution. Within an era, one of the length scales decreases monotonically to the past while the other two oscillate (see BKL, p. 544, Figure 2). When the universe enters a new era, the previously decreasing length scale will oscillate while one of the previously oscillating length scales will decrease monotonically. The Kasner epochs are separated by much shorter time intervals in which the model is approximated by a Taub vacuum Bianchi II solution (first discovered by Taub (1951); see Kramer et al (1980), p. 136). This pattern of evolution can, however, be interrupted by periods of anomalous behaviour, during which the model is approximated by a vacuum Bianchi VI₀ or VII₀ solution (referred to as the “case of small oscillations” by BKL, p. 539, also referred to as a “long era” by some authors), although this anomalous behaviour becomes increasingly less likely as the singularity is approached (BKL, p. 550). The evolution of the models in the case of small oscillations was investigated in detail by Khalatnikov and Pokrovsky (1972).

In the Hamiltonian approach (Misner (1969)) the vacuum Einstein field equations were reduced to a time-dependent Hamiltonian system for a particle, “the universe point”, in two dimensions. The system was analyzed heuristically by approximating the time-dependent potentials by moving “potential walls”. The Kasner epochs of BKL correspond to free motion of the universe point, the Taub vacuum transitions occur when the universe point reflects off the potential walls (“a bounce”), and the anomalous behaviour corresponds to motion in the “corner channels” of the potential walls (also referred to as a “mixing bounce”). A simple overview of this approach is given by Misner (1970) and by Barrow (1982). There has been some confusion in the literature about the interpretation of the BKL eras within the Hamiltonian approach. The potential walls for type VIII and IX models are roughly triangular in shape (e.g. Ryan and Shepley (1975), p. 194). Within an era (or cycle) the moving universe point is successively reflected by two sides of the triangles, corresponding to two length

scales undergoing oscillations. The next era commences when the universe point is successively reflected by a different pair of sides of the triangles. This permutation of the axes of oscillation, however, does not correspond to motion in the corner channels, which, as mentioned above, temporarily disrupts the sequence of Kasner epochs and eras (compare MacCallum (1979), p. 53, and Barrow (1982), p. 31).

In the dynamical systems approach (Bogoyavlensky and Novikov (1973)) the evolution of the universe model is described by an orbit of a system of differential equations in \mathbb{R}^5 . The sequence of Kasner epochs and Taub transitions is described in terms of an approximation of the orbits by sequences of separatrices (that is, orbits which join two saddle equilibrium points) which lie on an attractor, but the eras and the anomalous behaviour of BKL are not described. A comprehensive discussion of this method is given in Bogoyavlensky (1985).

Both the BKL and the Hamiltonian approach have the drawback that the asymptotic state near the big-bang is not well-defined mathematically, since the basic variables (the metric components and their time derivatives, relative to a time-independent invariant frame) diverge as the singularity is approached. On the other hand, the dynamical systems approach, while not suffering from this drawback, has unfortunately had relatively little impact, partly due to the use of variables that do not have a clear physical interpretation, and partly due to the fact that the attractor has a rather complicated structure. In this paper we overcome these difficulties by using expansion-normalized variables that are associated with an invariant orthonormal frame.

3. EXPANSION-NORMALIZED VARIABLES

We give a brief description of the basic variables that are used in our approach, referring to WH for full details and for the explicit form of the differential equations. We introduce a group invariant orthonormal frame $\{\mathbf{e}_0 = \frac{\partial}{\partial t}, \mathbf{e}_\alpha\}$, $\alpha = 1, 2, 3$, and use the Ellis-MacCallum (1969) formalism for the orthogonal Bianchi models, in which the basic variables are the non-zero commutation functions associated with this frame. For models of class A, which includes Bianchi types VIII and IX, the frame can be chosen so that

$$[\mathbf{e}_0, \mathbf{e}_\alpha] = \theta_\alpha(t)\mathbf{e}_\alpha, \quad \alpha = 1, 2, 3, \quad (\text{no summation})$$

$$[\mathbf{e}_1, \mathbf{e}_2] = n_3(t)\mathbf{e}_3, \quad \text{cycle over } 1, 2, 3.$$

The $\theta_\alpha(t)$ are the eigenvalues of the expansion tensor, and $n_\alpha(t)$ algebraically determine the spatial curvature (i.e. the curvature of the group orbits), and also the Bianchi type of the isometry group. It is convenient to replace the θ_α by the expansion scalar θ and the shear variables σ_\pm , defined by

$$\sigma_+ = \frac{3}{2}(\theta_2 + \theta_3 - \frac{2}{3}\theta), \quad \sigma_- = \frac{\sqrt{3}}{2}(\theta_2 - \theta_3).$$

The vector

$$\mathbf{x} = (\theta, \sigma_+, \sigma_-, n_\alpha) \in \mathbb{R}^6$$

is regarded as defining the *physical state* of the model at time t . The Einstein field equations with perfect fluid source and equation of state

$$p = (\gamma - 1)\mu$$

determine the evolution of \mathbf{x} through an autonomous differential equation

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}), \quad (1)$$

where $f : \mathbb{R}^6 \rightarrow \mathbb{R}^6$ is a polynomial function, homogeneous of degree two (see WH). The matter density μ is an auxiliary variable, and is a quadratic function in \mathbf{x} .

This formulation of the field equations is not suitable for investigating the behaviour near the big-bang $t = 0$, since the variable \mathbf{x} typically diverges as $t \rightarrow 0$. We thus introduce *expansion-normalized* variables, which are dimensionless variables defined by normalizing with the expansion scalar θ :

$$\Sigma_{\pm} = \frac{\sigma_{\pm}}{\theta}, \quad N_{\alpha} = \frac{n_{\alpha}}{\theta}$$

These are referred to as the *dimensionless shear variables* (Σ_{\pm}) and the *dimensionless spatial curvature variables* (N_{α}). On introducing a new time-variable τ , the differential equation (1) assumes the form

$$\frac{d\mathbf{X}}{d\tau} = F(\mathbf{X}), \quad (2)$$

where

$$\mathbf{X} = (\Sigma_+, \Sigma_-, N_{\alpha}) \in \mathbb{R}^5 \quad (3)$$

is regarded as defining the *dynamical state* of the cosmological model at time t (see WH, equation (2.24)). The function $F : \mathbb{R}^5 \rightarrow \mathbb{R}^5$ is a polynomial function (of degree 3) and hence is a smooth function on \mathbb{R}^5 . The big-bang $t = 0$ corresponds to $\tau \rightarrow -\infty$. The physical density is replaced by the dimensionless *density parameter*

$$\Omega = \frac{3\mu}{\theta^2}$$

which is quadratic in \mathbf{X} (see WH, equations (2.18), (2.27) and (2.28)). The differential equation for θ (part of the system (1)) decouples, accounting for the decrease in dimension from six to five. This equation, together with the equation which defines the new time variable τ , are the transition equations for passing from the dynamical state \mathbf{X} to the physical state \mathbf{x} :

$$\frac{d\theta}{d\tau} = -[1 + q(\mathbf{X})]\theta, \quad \frac{dt}{d\tau} = \frac{3}{\theta}.$$

Here $q(\mathbf{X})$ is the deceleration parameter, and is quadratic in \mathbf{X} (see WH, equation (2.26)).

The differential equation (2) for the expansion-normalized variables (3) form the basis for our analysis of the oscillatory singularity.

4. THE KASNER RING AND TAUB SEPARATRICES

In this section we describe the Kasner and Taub vacuum solutions (which played an important role in previous discussions of the oscillatory singularity) in terms of the expansion-normalized variables (Σ_{\pm}, N_{α}).

The one-parameter family of Kasner vacuum solutions corresponds to a circle \mathcal{K} of equilibrium points of the differential equation (2), given by

$$\Sigma_+^2 + \Sigma_-^2 = 1, \quad N_1 = N_2 = N_3 = 0,$$

with Σ_+ and Σ_- constant. We shall refer to this circle as the *Kasner ring*. Analysis of the eigenvalues shows that apart from three exceptional degenerate points, the equilibrium points of \mathcal{K} are saddles, and *into the past*, they attract in 3 dimensions, repel in one dimension and are neutral in one dimension, namely in the direction tangent to \mathcal{K} (see WH). The three exceptional points correspond to flat spacetime.

The Taub vacuum Bianchi II solutions can also be solved for explicitly. There are six equivalent families of solutions, given by

$$\Omega = 0, \quad N_1 > 0, \quad \text{or} \quad N_1 < 0, \quad N_2 = N_3 = 0,$$

with the remaining four families obtained by cycling on 1, 2, 3. The orbits are nonsingular, and are positively and negatively asymptotic to Kasner equilibrium points. We shall refer to them as *Taub separatrices*. It follows from the DE (2) that the orbits with $N_1 \neq 0$ are given, in the 3D subspace $N_2 = N_3 = 0$, by the intersection of the family of

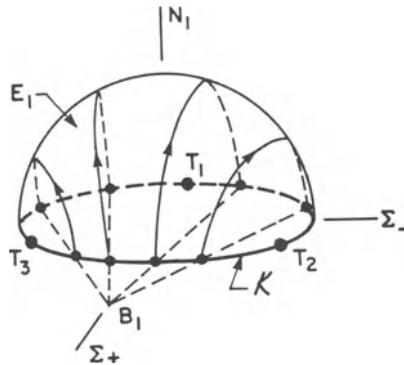


Figure 1. The Taub separatrices with $N_1 > 0$, $N_2 = N_3 = 0$, which lie on the upper half of the ellipsoid E_1 , given by $\Sigma_+^2 + \Sigma_-^2 + \frac{3}{4}N_1^2 = 1$, $N_2 = 0 = N_3$. The arrows indicate evolution into the past.

planes

$$\Sigma_- = k(\Sigma_+ - 2),$$

$k = \text{constant}$, with the vacuum surface $\Omega = 0$, which in this case is the ellipsoid

$$\Sigma_+^2 + \Sigma_-^2 + \frac{3}{4}N_1^2 = 1 \tag{4}$$

(see Figure 1). The projection of these orbits into the $\Sigma_+\Sigma_-$ -plane are straight lines which intersect at the point $(2,0)$ when extended outside the Kasner ring (see Figure 2). The other two families of Taub separatrices are obtained by rotating the first family successively through 120° in the $\Sigma_+\Sigma_-$ -plane (see the discussion of symmetry in WH).

Each non-exceptional Kasner equilibrium point P (i.e., all points on \mathcal{K} except for T_1, T_2, T_3 in Figure 2) has a 1D unstable manifold (*into the past*), which is formed by the two Taub separatrices which emanate (*into the past*) from P , one on the upper half-ellipsoid, and one on the lower half-ellipsoid. If P is on the arc T_2T_3 , for example, then

these unstable Taub separatrices are defined by the straight line which passes through the points P and B_1 . This line has a second point of intersection P' with \mathcal{K} , and hence can be thought of as mapping P onto P' . In this way, the Taub separatrices define a map B of the Kasner ring onto itself. Thus, given a point P_0 on \mathcal{K} , the successive action of this map determines a unique sequence of Kasner equilibrium points, defined by

$$P_{i+1} = B(P_i), \quad i = 0, 1, 2, \dots,$$

as illustrated in Figure 3. This map, which appeared in terms of different variables in Bogoyavlensky (1985) (see page 57), will be referred to as the *B-map*.

To conclude this section, we note that Collins (1971) gave a phase plane analysis of the Taub vacuum solutions. The resulting phase portrait is similar to our Figure 2. Collins' variables β'_1 and β'_2 are closely related to our Σ_{\pm} . It should be kept in mind

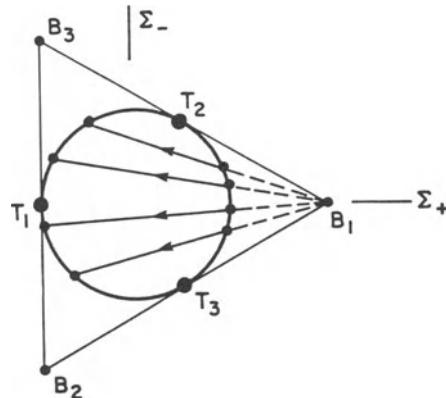


Figure 2. The projections of the Taub separatrices with $N_1 \neq 0$, $N_2 = N_3 = 0$, into the $\Sigma_+\Sigma_-$ -plane. The arrows indicate evolution into the past.

however, that our Figure 2 is the result of projecting down from \mathbb{R}^5 into the $\Sigma_+\Sigma_-$ -plane, while Collins' diagram is the phase portrait for a two-dimensional system.

5. ASYMPTOTIC BEHAVIOUR OF THE ORBITS

In WH, it was shown that differential equation (2) has no periodic orbits or recurrent orbits, due to the existence of certain functions which are monotone along the orbits. Thus the system should be regarded as dissipative, and one expects that the orbits, if bounded into the past, will be asymptotic to an attractor of dimension less than the dimension of the phase space. On the other hand, the system has no equilibrium points which are sinks, as regards evolution into the past. Thus the attractor must be more complicated than an equilibrium point or limit cycle.

A series of numerical experiments have been conducted for the DE (2) (see Ma (1988) for further details). They show that typical orbits of Bianchi types VIII and IX (i.e. with $N_1 N_2 N_3 \neq 0$) pass through a *transient stage*, at the end of which they

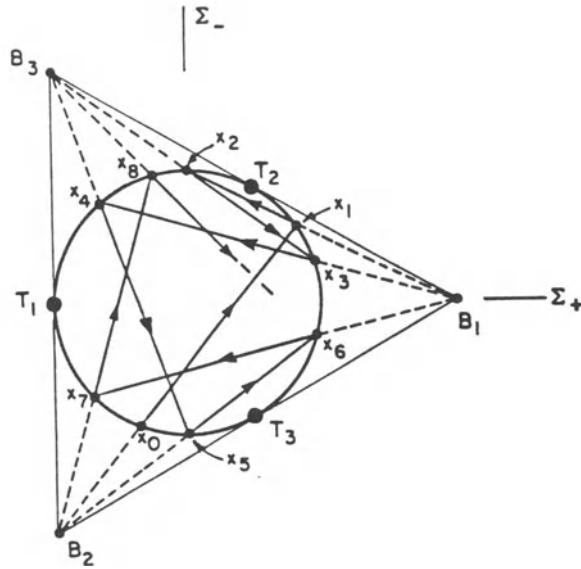


Figure 3. The action of the B-map on the Kasner ring, defining a sequence of Kasner equilibrium points.

approach the Kasner ring \mathcal{K} (into the past). Since the equilibrium points of \mathcal{K} are saddles, the orbits must subsequently leave \mathcal{K} , and this is confirmed by the experiments. In addition, the direction of departure is determined by the unique Taub separatrix which leaves the equilibrium point of \mathcal{K} . The numerical experiments show that the orbit follows this Taub separatrix until it again approaches the Kasner ring. The orbit again departs, following the appropriate Taub separatrix, and this process continues indefinitely. In other words, during the *asymptotic stage*, the orbit is approximated by an infinite sequence of Taub separatrices joining Kasner equilibrium points. Thus, *the projection of the orbit into the $\Sigma_+\Sigma_-$ -plane is determined by the action of the B-map on the Kasner ring*, as in Figure 3. This means that on completion of the transient stage, the result of a numerical experiment for the DE (2) can be constructed graphically using a unit circle, ruler and pencil!

The point X , which represents the dynamical state of the universe, does not move uniformly along this separatrix sequence, however, but spends most of the time (τ -time) near the Kasner points, since they are equilibrium points. This resting on the Kasner ring corresponds to the Kasner epochs of BKL. The rapid transition of a Taub separatrix corresponds to a “bounce” in the Hamiltonian approach. In our approach, the statement that “the universe is approximated by a Kasner model” means that the universe point $X = (\Sigma_+, \Sigma_-, N_\alpha)$ is close to a Kasner equilibrium point in the dynamical phase space, i.e. that the dimensionless shear and spatial curvature variables are close to the corresponding Kasner values.

The qualitative behaviour of the orbits during the transient stage, as determined numerically, can be explained heuristically as follows. The function

$$\Delta_1 = (N_1 N_2 N_3)^{2/3}$$

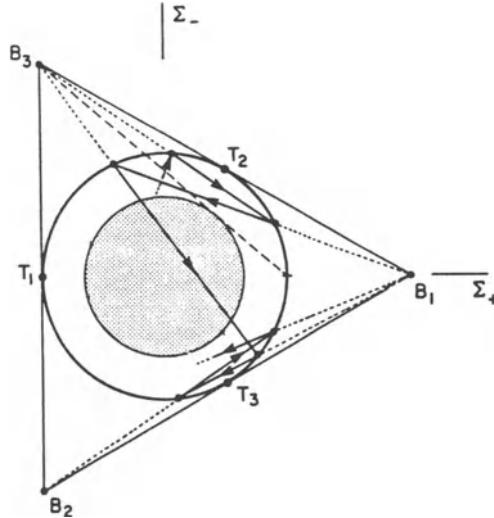


Figure 4. Projection of a separatrix sequence in the $\Sigma_+\Sigma_-$ -plane, showing a change of era.

is monotone decreasing into the past along any type VIII or IX orbit (see WH). It follows that the α -limit set of such an orbit belongs to the subset $\Delta_1 = 0$ (assuming that the limit set is non-empty). Thus we expect that as $\tau \rightarrow -\infty$, a type VIII or IX orbit will be approximated by orbits in the subset $\Delta_1 = 0$, i.e. by orbits of more special Bianchi types. Since the α -limit set of almost all orbits in the subset $\Delta_1 = 0$ is an equilibrium point on the Kasner ring, one expects that almost all type VIII or IX orbits will likewise approach the Kasner ring, and then follow a separatrix sequence. We hope to be able to make this argument precise, and details will be published elsewhere.

The qualitative behaviour of the orbits in the asymptotic stage enables us to give a simple description of the BKL eras, as follows. A BKL era corresponds to a part of a separatrix sequence which contains lines emanating from only two of the three points B_1 , B_2 , and B_3 in Figure 3. By studying Figure 3, one concludes that a change of era occurs when the separatrix sequence passes sufficiently close to the centre of the Kasner ring. A change of era is illustrated in Figure 4, and occurs when the projection of a Taub separatrix passes through the shaded disk. It can be deduced from the geometry of the B-map that the radius of this disk is $\sqrt{\frac{3}{7}}$ (Ma (1988)). This leads to a simple interpretation of the eras in terms of the dimensionless shear parameter:

$$\Sigma = \Sigma_+^2 + \Sigma_-^2$$

(see WH, equations (2.16) and (2.27)). During each Taub transition within a particular era, the shear parameter assumes successively smaller values until it attains a value less than $\frac{3}{7}$, which signals the start of a new era. This is illustrated in Figure 5, in which the output from a numerical experiment is displayed.

6. THE ATTRACTOR

Various definitions of attractor have been given in the literature. The definition of Milnor (1985) is the one which is most appropriate for our purposes. Essentially, the (past) attractor is the smallest closed subset that contains the α -limit sets of typical orbits. The numerical experiments suggest that the union of the Kasner ring and the Taub separatrices is the (past) *attractor* for the DE (2), as $\tau \rightarrow -\infty$. This set, denoted A , is the union of the three ellipsoids E_1, E_2, E_3 referred to in section 4:

$$A = E_1 \cup E_2 \cup E_3.$$

These ellipsoids are defined by

$$E_1 : \Sigma_+^2 + \Sigma_-^2 + \frac{3}{4}N_1^2 = 1, \quad N_2 = N_3 = 0,$$

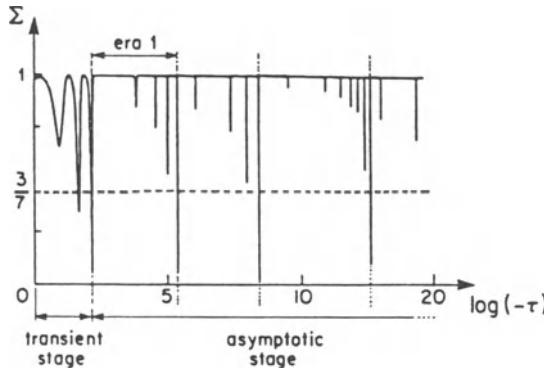


Figure 5. The shear parameter Σ as a function of $\log(-\tau)$, as $\tau \rightarrow -\infty$, showing the BKL eras.

with E_2, E_3 obtained by cycling on 1, 2, 3. Their intersection is the Kasner ring \mathcal{K} :

$$E_1 \cap E_2 \cap E_3 = \mathcal{K}.$$

Some analytical evidence that the set A is an attractor can be obtained by considering the density parameter Ω and the scalar Δ_2 , which is defined by

$$\Delta_2 = [(N_1 N_2)^2 + (N_2 N_3)^2 + (N_3 N_1)^2]^{1/2}.$$

The attractor is then given by

$$\Delta_2 = 0, \quad \Omega = 0.$$

Thus in order to prove that the set A is an attractor, it is necessary to prove that

$$\lim_{\tau \rightarrow -\infty} \Delta_2 = 0, \quad \lim_{\tau \rightarrow -\infty} \Omega = 0 \tag{5}$$

for almost all orbits. Note that $\Delta_2 = 0$ implies that at least two of the N_α are zero, and hence that $\Delta_1 = 0$. We refer to WH, Figure 1, for the interpretation of Δ_1 and Δ_2 .

Unlike Δ_1 , the functions Δ_2 and Ω are not monotone along orbits, even in a neighbourhood of A . However, since the evolution equation for the density parameter is (here the prime is defined by $' = d/d\tau$):

$$\Omega' = [2q - (3\gamma - 2)]\Omega,$$

and $q = 2$ on \mathcal{K} (see WH), it follows that if $\gamma < 2$, then there is a neighbourhood U of \mathcal{K} such that $\Omega' > 0$ along all orbits in U for which $\Omega > 0$. This means that Ω *decreases* exponentially into the past along all orbits in U . In addition, it can be shown that each point of \mathcal{K} , except for the points T_α , has a neighbourhood \tilde{U} such that $\Delta_2' > 0$ along all non-singular orbits in \tilde{U} for which $\Delta_2 > 0$. This means that Δ_2 *decreases* exponentially into the past along all orbits in \tilde{U} . Since the universe point, while following a separatrix sequence, spends most of its time near the Kasner ring, it is plausible that Ω and Δ_2 will overall decrease exponentially, and hence tend to zero as $\tau \rightarrow -\infty$. Thus the challenge remains to prove the results (5), for almost all orbits.

It is of interest to note that the statement $\lim_{\tau \rightarrow -\infty} \Omega = 0$ for almost all orbits corresponds to the claim by BKL that the matter is dynamically negligible near the singularity (even though the matter density μ diverges). It is important to note that the limits (5) do not hold for all orbits, i.e. there are sets of orbits of dimension less than or equal to three for which $\lim_{\tau \rightarrow -\infty} \Omega \neq 0$ and/or $\lim_{\tau \rightarrow -\infty} \Delta_2 \neq 0$. These non-generic orbits were described in WH. They correspond to orbits whose α -limit set is an equilibrium point other than a Kasner equilibrium point.

We have seen that the asymptotic behaviour of Ω depends on the equation of state parameter satisfying $\gamma < 2$. The value $\gamma = 2$ is special, and a bifurcation occurs there. If $\gamma = 2$, the Kasner ring becomes a disk \mathcal{D} of equilibrium points, a triangular subset of which are sinks as regards evolution into the past (see WH). Numerical experiments suggest that the α -limit set of a typical orbit is an equilibrium point of this subset. Thus if $\gamma = 2$ the oscillatory behaviour does not occur. This conclusion has been reached by other authors (e.g. Barrow (1978), p. 212).

In order to discuss the anomalous behaviour of BKL, we require the concept of Lyapunov stability. Recall that an invariant set A of a dynamical system Φ on \mathbb{R}^n is said to be (positively) *Lyapunov stable* if for all neighbourhoods U of A there exists a neighbourhood V of A such that $\Phi(t, V) \subset U$ for all $t > 0$. Although the numerical experiments strongly suggest that the set A is an attractor, it is important to note that it is not a Lyapunov stable set. This is essentially due to the fact that every neighbourhood of the three exceptional equilibrium points T_α contains orbits along which $\Delta_2' < 0$. This implies that Δ_2 *increases into the past*, and hence that these orbits move away from the set A . This behaviour can be described qualitatively by considering the vacuum solutions of Bianchi types VI₀ and VII₀. It can be shown that vacuum orbits of types VI₀ and VII₀ with $N_1 = 0$ are past asymptotic to a Kasner equilibrium point in the arc $T_2 T_3$ in Figure 2, and are either (type VI₀) future asymptotic to the point T_1 , or (type VII₀) are future asymptotic to a point on the lines of equilibrium points which are denoted \mathcal{L}_1 in WH, and which join the point T_1 . The behaviour of the other families of type VI₀ and VII₀ orbits, with $N_2 = 0$ or $N_3 = 0$, is obtained by cycling on 1, 2, 3. Thus, if a generic orbit of type VIII or IX, while following a separatrix sequence on the attractor A (into the past), approaches too close to one of the exceptional points T_α , say T_1 , the orbit can temporarily leave the attractor, by following a vacuum VI₀

(if $N_2 N_3 < 0$) or a vacuum VII_0 orbit (if $N_2 N_3 > 0$), until it subsequently returns to the Kasner ring along the arc $T_2 T_3$. In summary, the *anomalous behaviour that can occur*, i.e. the “case of small oscillations” of BKL and the “corner channel motion” of the Hamiltonian approach, *manifests itself in the negative Lyapunov instability of the attractor A* , and is governed by the vacuum orbits of Bianchi types VI_0 and VII_0 .

We conclude this section by making some comparisons between the attractor A of the DE (2) and the attractor P of Bogoyavlensky (1985) (henceforth referred to as BY). The attractor P is the union of three circles of equilibrium points and various families of separatrices, of dimension one or two, which join points on these circles (BY, pp. 61 and 65). Thus P is a 3-dimensional set whereas A is of dimension 2. The 1-dimensional separatrices are interpreted in terms of the BKL oscillations, but the 2-dimensional separatrices are not given a physical interpretation (BY, p. 68). There is another significant difference. In the Bogoyavlensky approach, the boundary of the physical region of phase space consists of a physical part and an unphysical part (BY, p. 46). The attractor P lies in the unphysical part, and hence consists of unphysical orbits (BY, p. 65), whereas the attractor A is the union of physical orbits. For example, the three circles of equilibrium points in BY are unphysical, whereas the equilibrium points of our Kasner ring correspond to the Kasner solutions. Thus, when an orbit is close to the attractor A , the corresponding model is approximated in a well-defined way by the model whose orbits form the attractor. These differences originate in the initial choice of variables. We use the commutation functions associated with an *orthonormal* group-invariant frame, whereas BY uses the spatial metric components associated with a *time-independent* group-invariant frame. The unphysical part of the boundary in BY is where the spatial metric is degenerate.

7. CONCLUDING REMARKS

We have shown that the oscillatory approach to the singularity in orthogonal Bianchi models of types VIII and IX is generated by a two dimensional attractor of the cosmological dynamical system. The attractor is the union of three ellipsoids in \mathbb{R}^5 , and consists of the Kasner ring and the Taub separatrices. The Kasner epochs and eras, the Taub transitions and the anomalous behaviour are all described in terms of properties of the dynamical system and its attractor. Our approach also shows that the asymptotic structure of the orbits of the dynamical system (into the past) can be described by the iterates of a one dimensional map, the B-map, acting on the Kasner ring, giving a sequence of Kasner epochs.

The final issue that we wish to address is that of *chaotic behaviour*. The Kasner solutions form a 1-parameter family, and a standard choice of parameter is the BKL parameter $u \in [1, \infty)$, which determines the Kasner exponents p_α according to

$$p_1(u) = \frac{-u}{1+u+u^2}, \quad p_2(u) = \frac{1+u}{1+u+u^2}, \quad p_3(u) = \frac{u(1+u)}{1+u+u^2}$$

(BKL, p. 528). A change in Kasner epoch is described by

$$u \rightarrow u - 1, \quad \text{if } 2 < u$$

$$u \rightarrow \frac{1}{u-1}, \quad \text{if } 1 < u < 2 \tag{6}$$

(BKL, p. 537), the second case indicating the start of a new era.

It is convenient to describe these changes by means of the iterates of a map on a finite interval, and this can be done as follows. If one writes

$$u = k + x,$$

where k is a positive integer and $0 < x < 1$, then according to equation (6), during an era x remains constant, while k successively decreases by 1 until $k = 1$. In the first epoch of the next era, the BKL parameter is given by

$$\bar{u} = \frac{1}{u-1} = \frac{1}{x},$$

which can be written

$$\bar{u} = \bar{k} + \bar{x},$$

where

$$\bar{k} = \left[\frac{1}{x} \right]$$

is the number of epochs in the new era, and

$$\bar{x} = \frac{1}{x} - \left[\frac{1}{x} \right].$$

Here $[x]$ denotes the integer part of the enclosed real number, x . The change $x \rightarrow \bar{x}$ when a new era commences, is described by the map $T : [0, 1] \rightarrow [0, 1]$ defined by

$$T(x) = \frac{1}{x} - \left[\frac{1}{x} \right] \quad (7)$$

(Barrow (1982), Khalatnikov et al (1985)). The iterates of this map contain the information about the sequence of Kasner epochs and eras. Alternatively, one can replace the BKL parameter u by a new parameter $v \in [0, 1]$, such that

$$v = \frac{1}{u}.$$

The value $v = 0$, which corresponds to $u \rightarrow \infty$, describes the exceptional Kasner equilibrium points T_α . The changes in epoch, as governed by equation (6), are then described by the iterates of the map $G : [0, 1] \rightarrow [0, 1]$, defined by

$$G(v) = \begin{cases} \frac{v}{1-v}, & 0 \leq v \leq \frac{1}{2} \\ \frac{1-v}{v}, & \frac{1}{2} < v \leq 1 \end{cases} \quad (8)$$

The map G is essentially an equivalent representation of the B-map on the Kasner ring (Ma (1988), p. 41).

It has been shown that the maps T and G are chaotic (Barrow (1982), Ma (1988)). It is thus natural to ask whether the DE (2), which gives rise to these maps, is chaotic. The attractor of the DE is a two dimensional set, the union of three ellipsoids in \mathbb{R}^5 . In addition, the orbits which it contains are either equilibrium points (the Kasner ring) or are non-singular orbits (the Taub separatrices) which are past and future asymptotic to equilibrium points. It follows that the attractor is neither strange nor chaotic (see Milnor (1985), and Grebogi et al (1984) for this terminology). It thus appears that the DE (2) is not chaotic. Another criterion for the occurrence of chaos is the existence of

a positive Lyapunov exponent (e.g. Shimada and Nagashima (1979)). Numerical work by Zardecki (1983), using the BKL equations, suggested that there exists a positive Lyapunov exponent. However, more recent work by Francisco and Matsas (1988), Burd et al (1989), and Hobill et al (1989), using both the BKL equations and the DE (2), casts doubt on Zardecki's result and suggests that there are no positive exponents, which supports our conclusion based on the study of the attractor. Thus it appears that the five-dimensional cosmological dynamical system, defined by the DE (2), is not chaotic as regards its attractor and Lyapunov exponents, and yet it has a one dimensional discretization, defined by the map (7) or (8), which is chaotic.

ACKNOWLEDGEMENTS

One of the authors (JW) gratefully acknowledges the hospitality of Zoltan Perj  s and his colleagues at the Third Hungarian Relativity Workshop, and the financial support of the Hungarian Academy of Sciences. We would like to thank C.B. Collins, C.G. Hewitt and L. Hsu for helpful comments and discussions and A. Burd and D. Hobill for discussing their results prior to publication.

This manuscript was typeset in \LaTeX by Teviet Creighton.

REFERENCES

- Anosov, D.V. and Arnold, V.I. (1988), *Dynamical Systems I*, Springer-Verlag, New York, NY.
- Barrow, J.D. (1979), *Nature*, **272**, 211.
- Barrow, J.D. (1982), *Phys. Rep.*, **85**, 1.
- Belinskii, V.A. and Khalatnikov, I.M. (1969), *Sov. Phys. JETP*, **29**, 911.
- Belinskii, V.A., Khalatnikov, I.M. and Lifshitz, E.M., (1970), *Adv. Phys.*, **19**, 525.
- Bogoyavlensky, O.I., and Novikov, S.P. (1973), *Sov. Phys. JETP*, **37**, 747.
- Bogoyavlensky, O.I. (1985), *Methods in the Qualitative Theory of Dynamical Systems in Astrophysics and Gas Dynamics*, Springer Verlag, New York, NY.
- Burd, A.B., Buric, N. and Ellis, G.F.R. (1990), *Gen. Rel. Grav.*, **22**, 349.
- Ellis, G.F.R. and MacCallum, M.A.H. (1969), *Commun. Math. Phys.*, **12**, 108–141.
- Francisco, G. and Matsas, G.E.A. (1988), *Gen. Rel. Grav.*, **20**, 1047.
- Grebogi, C., Ott, E., Pelikan, S., and Yorke, J.A. (1984), *Physica D*, **13**, 251.
- Hobill, D., Bernstein, D., Simkins, D. and Welge, M. (1989), in *Proc. 12th Int. Conf. Gen. Rel. Grav.*, Boulder, Univ. of Colorado, abstracts p. 337.
- Khalatnikov, I.M. and Pokrovsky, V.L. (1972), in *Magic without Magic*, (ed. J. Klauder), W. H. Freeman, San Francisco, CA.
- Khalatnikov, I.M., Lifshitz, E.M., Khanin, K.M., Shchur, L.N., and Sinai, Ya. G. (1985), *J. Stat. Phys.*, **38**, 97.
- Kramer, D., Stephani, H., MacCallum, M.A.H. and Herlt, E. (1980), *Exact Solutions of Einstein's Field Equations*, Deutscher Verlag der Wissenschaften, Berlin, and Cambridge University Press, Cambridge, UK.

- Ma, P.K-H. (1988), *A Dynamical Systems Approach to the Oscillatory Singularity in Cosmology*, M. Math. thesis, University of Waterloo.
- MacCallum, M.A.H. (1979), in *Physics of the Expanding Universe*, (ed. M. Demianski), *Lecture Notes in Physics*, Vol. 109, Springer-Verlag, Berlin and Heidelberg.
- Milnor, J. (1985), *Commun. Math. Phys.*, **99**, 177.
- Misner, C.W. (1969), *Phys. Rev. Lett.*, **22**, 1071.
- Misner, C.W. (1970), in *Relativity*, (eds., M. Carmelli, S. Fickler, and L. Witten), Plenum Publishing, New York, NY.
- Nemytskii, V.V. and Stepanov, V.V. (1960), *Qualitative Theory of Differential Equations*, Princeton University Press, Princeton, NY.
- Ryan, M. P., Jr. and Shepley, L.C. (1975), *Homogeneous Relativistic Cosmologies*, Princeton Univ. Press, Princeton, NJ.
- Shimada, I. and Nagashima, T. (1979), *Prog. Theoret. Phys.*, **61**, 1605.
- Taub, A. (1951), *Ann. Math.*, **53**, 472.
- Wainwright, J. and Hsu, L. (1989), *Class. Quantum Grav.*, **6**, 1409.
- Zardecki, A. (1983), *Phys. Rev.*, **D28**, 1235.

LIST OF PARTICIPANTS

Hossein Abolghasem

Dept. of Mathematics, Statistics, and Computer Science, Dalhousie University
Halifax, Nova Scotia, B3H 3J5 Canada

houssein@cs.dal.ca

Beverly Berger

Dept. of Physics, Oakland University
Rochester, MI 48309 USA
berger@vela.acs.oakland.edu

Vikas Bhushan

Dept. of Physics, University of British Columbia
6224 Agricultural Road, Vancouver, BC, V6T 2A6 Canada
vikas@physics.ubc.ca

Luca Bombelli

RGGR, Universite Libre de Bruxelles
Campus Plaine CP 231, 1050 Bruxelles, Belgium
lbombell@ulb.ac.be

Adrian Burd

Dept. of Oceanography, Dalhousie University
Halifax, Nova Scotia, B3H 4J1 Canada
aburd@open.dal.ca

Esteban Calzetta

Instituto de Astronomica y Fisica del Espacio
Casilla de Correo 67 - Sucursal 28, 1428 Buenos Aires, Argentina
(calzetta@iafe.edu.ar)

Matthew Choptuik
Center for Relativity, Department of Physics
University of Texas, Austin, TX, 78712-1081 USA
matt@einstein.ph.utexas.edu

Piotr Chruściel
Max Planck Institut für Astrophysik
1, Schwarzschild Straße, D 85740 Garching bei München, Germany
piotr@mpa-garching.mpg.de

Nyuyen Hong Chuong
Hanoi Institute of Theoretical Physics
P. O. Box 429 Boho, 10000 Hanoi, Vietnam
(Department of Physics, Syracuse University, Syracuse, NY 13244-1130, USA)
chuong@suhep.phy.syr.edu

Richard Churchill
Dept. of Mathematics, Hunter College and Graduate School, CUNY
695 Park Avenue, New York, NY, 10021 USA
richc@cunyvm.cuny.edu

Alan Coley
Dept. of Mathematics, Statistics, and Computer Science, Dalhousie University
Halifax, Nova Scotia, B3H 3J5 Canada
aac@cs.dal.ca

George Contopoulos
Dept. of Astronomy, University of Athens
Panepistimiopolis, GR 15783 Zografos, Athens, Greece
(Dept. of Astronomy, University of Florida, Gainesville, FL 32611 USA)
spm11@grathun1.bitnet (gco@astro.ufl.edu)

W. Eugene Couch
Dept. of Mathematics, University of Calgary
Calgary, AB, T2N 1N4 Canada
couch@acs.ucalgary.ca

Teviet Creighton
Dept. of Physics and Astronomy, University of Calgary
Calgary, AB, T2N 1N4 Canada
tdcreigh@acs.ucalgary.ca

Richard Cushman
Mathematics Institute, University of Utrecht
3508TA Utrecht, The Netherlands
cushman@math.ruu.nl

Bahman Darian
Department of Physics, University of Alberta
Edmonton, AB, T6G 2J1 Canada
darian@phys.ualberta.ca

George F. R. Ellis
Department of Applied Mathematics, University of Cape Town
Rondebosch, C.P. 7700, South Africa
ellis@appmath.uct.ac.za

Gerson Francisco
Instituto de Fisica Teorica (UNESP)
Rua Pamplona 145, 01405-900-São Paulo, São Paulo, Brazil
gf_akham@ift.uesp.ansp.br

Alistair Fraser
Dept. of Meteorology, Pennsylvania State University
University Park, PA, 16802 USA
abf1@psuvm.psu.edu

David Hobill
Dept. of Physics and Astronomy, University of Calgary
Calgary, AB, T2N 1N4 Canada
hobill@scri.crag.ucalgary.ca

Malcolm MacCallum
School of Mathematical Sciences, Queen Mary and Westfield College
University of London, Mile End Road, London E1 4NS, U.K.
mm@maths.qmw.ac.uk

George Matsas
Instituto de Fisica Teorica (UNESP)
Rua Pamplona 145, 01405-900-São Paulo, São Paulo, Brazil
matsas@ift.uesp.ansp.br

Charles W. Misner
Dept. of Physics, University of Maryland
College Park, MD 20740 USA
misner@umdhep.umd.edu

Marcelo Ribeiro
Departamento de Astrofísica, Observatorio Nacional, CNPq
rua General Jose Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil
mbr@obsn.on.br

Svend Rugh
Niels Bohr Institute

Blegdamsvej 17, 2100 Copenhagen O, Denmark
rugh@nbivax.nbi.dk

Jędrzej Śniatycki
Dept. of Mathematics, University of Calgary
Calgary, AB, T2N 1N4 Canada
sniatyck@acs.ucalgary.ca

Reza Tavakol
School of Mathematical Sciences, Queen Mary and Westfield College
University of London, Mile End Road, London E1 4NS, U.K.
reza@maths.qmw.ac.uk

Roman Tomaschitz
Physics Dept., University of the Witwatersrand,
WITS 2050, Johannesburg, SA.
005ROMAN@witsvma.wits.ac.za

Robert Torrence
Dept. of Mathematics, University of Calgary
Calgary, AB, T2N 1N4 Canada
torrence@acs.ucalgary.ca

Robert van den Hoogen
Dept. of Mathematics, Statistics, and Computer Science, Dalhousie University
Halifax, Nova Scotia, Canada B3H 3J5
hoogen@cs.dal.ca

John Wainwright
Department of Applied Mathematics, University of Waterloo
200 University Avenue W., Waterloo, Ontario N2L 3G1, Canada
jwainwri@math.uwaterloo.ca

INDEX

- Anosov flow, 120
Arnold Cat Map, 207, 242
attractor, 62, 117, 193, 272
 Lorenz, 120, 392, 400
 strange, 62, 118, 126, 272, 294
- B map, 442–450, 456, 457
 periodic solutions, 447
- B-map, 15
- Baker's Transformation, 209
- Bernoulli property, 379
- Bianchi IX, 5–16, 206, 215, 239, 331, 347, 348, 357, 425, 435
 3-volume, 9, 369
 as geodesic flow, *see* Bianchi IX,
 geodesic flow
 attractor, 456, 459–461
- BKL approximation, *see* BKL ap-
proximation
- Bogoyavlensky formalism, 399, 461
- chaos, 224, 344, 348, 350, 359, 393, 426, 461–463
 gauge invariance, 350, 383–400
 source of, 332, 338, 341, 379
- Chitre's variables, 323, 351, 409
- constants of motion, 344, 368
- coordinate systems, 389–390
- dynamical systems approach, 453
- Ellis-MacCallum-Wainwright formal-
ism, 13–16, 436–438, 451, 453–
454
- ergodic, 389, 414
- Gauss map, *see* Gauss map
- geodesic flow, 323–330, 351, 408–413
 problems, 326–330
- Hamiltonian, 10, 11, 320, 334
- Hamiltonian formalism, 10–13, 320–
321, 452
- intermittency, 396
- Kasner model, 332
- Kasner model, and, 7, 321, 336, 342, 348, 351, 358
- Kasner ring, 14, 439, 443
- Kolmogorov entropy, 349
- Lyapunov exponents, 8, 9, 332, 338, 348–351, 358–359, 383–400, 426
 choice of time, 10, 339, 358, 365, 385, 386, 433
- measures of chaos, 348–354
 problems with, 332, 383, 414
- mixing, 241, 425
- Mixmaster, 5–16, 188, 319, 331, 332, 333, 342, 357, 366–383
 anomalous behaviour, 342
 chaotic, 344
 constants of motion, 433
 era-epoch map, 9, 332, 336
 history, 322–323
 integrable, 426, 431
 Kasner behaviour, *see* Bianchi IX,
 Kasner model
 metric oscillations, 335
 physical relevance, 4, 335
 quantised, 407, 413
 structural stability, 401

numerical solution, 9, 334, 342, 349, 350, 374, 438–442
 oscillatory singularity, *see* singularity, Mixmaster potential, 320
 SDIC, 335, 338
 symbolic dynamics, 390–392
 tidal field, 378
 time variable, 332, 333, 339, 351, 394, 426, 433, 437
 time variable, and, 6, 11, 109, 374
 Bianchi models, 186–193, 215, 364
 dynamical systems, 191
 Hamiltonian formulation, 190
 self similar, 299
 bifurcation, 57–59
 Hopf, 58
 period doubling, 131, 139
 tangent, 139
 universal ratio, 139
 BKL approximation, 5, 331–334, 336, 342, 344, 349, 442
 accuracy, 9, 342, 344, 377, 436, 439, 457, 460
 anomalous behaviour, 460
 Farey map, 380
 Lyapunov exponent, 338, 358
 BKL map, *see* BKL approximation
 black hole
 and chaos, 149–156
 chaotic geodesics, 3, 131–145, 354
 homoclinic chaos, 212
 Melnikov method, 211
 null, 133–136
 timelike, 136–145, 211–212
 evaporation, 176
 formation, 3
 critical behaviour, 166–173, 175, 176
 mass scaling, 172
 numerical, 161
 in FLRW universe, 194, 195
 Newtonian orbits, 142
 null geodesics, 131, 133
 periodic orbits, 137
 perturbed, 152
 perturbed Schwarzschild, 208
 primordial, 194
 quasi-periodic geodesics, 137
 Schwarzchild, 150
 string capture, 149
 timelike geodesics, 136, 148, 150–155
 Melnikov method, 153
 two fixed, 132, 149
 bounce map, 339–342
 bulk viscosity, 225, 227, 306
 Cantor set, 62, 69, 70, 118, 131, 135, 142, 256
 fractal dimension, 70
 self similarity, 70
 Cauchy horizon, 115, 116
 chaos, 62, 139, 169, 361, 393, 394
 algebraic, 366
 definition, 88, 109–113, 205
 discrete and continuous systems, 349
 dissipation and, 230
 establishing existence, 75, 84, 85, 148
 FLRW models, *see* FLRW model, chaos
 galaxy distribution, 254
 gauge invariance, 109, 149, 220, 232, 240, 399
 generic, 206, 224, 400
 Hamiltonian systems, *see* Hamiltonian systems
 homoclinic, *see* homoclinic chaos
 integrability, 105
 Lyapunov exponent, 338, 392
 meaning of, 354
 metrical, 366
 quantum cosmology, *see* quantum cosmology, chaos
 relevance, 223
 SDIC, 69, 75, 88
 dangers, 113
 views on, 227
 chaotic cosmology programme, 416
 chaotic flow, 110
 chaotic foliation, 110
 chaotic scattering, 135, 139
 Chitre's variables, *see* Bianchi IX
 coarse graining, 230
 compactification scale, 245
 complexity, 395
 conformal time, 4, 216

- constants of the motion, 73
 Bianchi IX, 344, 368
 cosmological constant, 224, 306
 critical behaviour, *see* black hole, formation

 de Sitter spacetime, 186, 225, 227, 306
 de Vaucouleurs power law, 275, 276
 density parameter, 186, 241, 437
 dynamical system, 65, 66, 72, 73
 chaos
 fractal, 70
 chaotic, 69, 75
 cosmological models, 181, 189, 191
 equilibrium point, 21, 25, 28, 29, 32
 Hartman-Grobman Theorem, 46
 nonlinear, 47, 48
 stability, 21, 32
 group actions, 78
 Hamiltonian system, 36–38
 hyperbolic flow, 32
 Julia and Mandelbrot sets, 71
 limit set, 50
 linearization, 22, 45
 orbit, 21, 25
 periodic, 25, 48
 periodic orbit, 25
 smooth, 73
 stable manifold, 46, 85
 topological equivalence, 31
 unstable manifold, 46, 85
 dynamical systems
 cosmological models, 3

 Einstein-de Sitter model, 186, 286
 hierarchical, 287
 inhomogeneity, 287, 288
 observational relations, 286
 Einstein-Strauss geometry, 282
 energy condition, 116, 118
 violation, 350, 370
 equation of state, 186, 302, 460
 barotropic, 307, 368, 438
 dimensionless, 303, 306
 stiff matter, 370
 era scale parameter, 341
 ergodic, 83, 117, 231, 259, 351, 425
 expansions normalized variables, *see*
 Bianchi IX, Ellis-MacCallum-Wainwright formalism

 Farey map, 365, 377, 379, 403
 bold, 379
 Farey tree, 380
 fixed point, 66, 75, 117, 212
 dynamical system, *see* dynamical system, equilibrium point
 Gauss map, 344
 flow, 72, 73, 75, 78, 348, 350
 geodesic, *see* geodesic flow
 FLRW model, 3, 182, 185–186, 273, 286, 305
 $k = +1, 5, 215, 240$
 Melnikov integral, 219
 $k = -1, 239$
 chaos, 3, 208, 254
 geodesics
 $k = -1, 240$
 stretching effects, 242
 mixing and topology, 245
 particle creation, 266
 scalar field, 214–225
 chaotic, 217
 fractal, 70, 71, 256, 263, 274, 276, 280
 fractal dimension, 70, 272, 277, 278, 288, 291
 Lyapunov exponent, 294
 self similarity, 88
 Friedmann equation, 186, 302
 Friedmann-Robertson-Walker model, *see*
 FLRW model

 Gauss map, 9, 332, 338, 348, 349, 377, 379, 403
 fixed points, 344
 invariant probability, 338
 Kolmogorov entropy, 349
 general relativity
 chaos
 consequences, 148
 new canonical variables, 231
 non-linear nature, 2, 147
 geodesic flow, 80, 81–83, 211, 240, 353
 Bianchi IX, 351–353
 geodesics
 on 3-manifold, 254
 black hole, *see* black hole
 chaotic, 132, 149
 ergodic, 259
 multiply connected spaces, 259, 261

globally hyperbolic, 118
 Gowdy cosmology, 309–314
 asymptotic behaviour, 311

 Hénon-Heiles system, 86, 87
 Hamiltonian, 218
 Bianchi IX, 11, 320, 334, 352, 426, 432
 Bianchi models, 333
 meromorphic, 105
 Mixmaster bounces, 351
 perturbation, 153, 207, 208, 211
 Hamiltonian cosmology, 189, 214, 224
 homoclinic chaos, 214–225
 Hamiltonian system, 36–38, 72, 80, 112, 149, 206
 chaotic, 75, 83, 88, 223
 energy surfaces, 74
 numerical methods, 163, 220, 350
 reduction, 83, 227
 Hartman-Grobman Theorem, 45
 hierarchical cosmological models, 274–294
 homeomorphism, 31
 homoclinic chaos, 149, 205–214
 black hole geodesics, 212
 Hamiltonian cosmology, 214–225
 Melnikov method, 206, 210–211
 resonances, 212
 homoclinic orbit, 53, 149, 151, 354
 black hole, 150
 homoclinic point, 85, 87, 347
 transverse, 85
 homothety, 193, 300
 hyperbolic flow, 32

 inflation, 183, 186, 216, 217, 225, 383
 information, 149, 209, 361
 loss, 224, 358
 inhomogeneous cosmologies, 193–197
 integral curve, 73
 intermittency, 380, 396
 invariant measure, 380
 invariant set, 41, 62, 73, 208
 invariant tori, 60–62

 Jacobi equation, 84
 Jacobi metric, 81, 353
 Hénon-Heiles system, 87
 Jordan form, 23–24

 Julia set, 71
 k-flows, 240
 KAM curves, 142
 KAM theory, 139, 218
 KAM tori, 220, 221
 Kantowski-Sachs model, 187, 193
 Kasner epoch, 9, 16, 332, 357, 374, 378, 439, 443, 461
 Kasner era, 9, 332, 439, 458
 Kasner indices, 7, 334, 337, 340, 349, 439, 461
 and Kasner parameter, 334
 Kasner model, *see* Bianchi IX, Kasner model
 Kasner parameter, 332, 349, 358, 377
 Kasner ring, 454–456
 Kasner solution, 7, 333
 Kepler motion, 433
 Killing vector, 182, 300, 301, 311
 Bianchi models, 187
 Klein-Gordon equation, 4, 216, 304
 Kolmogorov entropy, 207, 209, 348
 Gauss map, 349

 lapse function, 10, 160, 333, 363
 large scale matter distribution, 273
 level surfaces, 73
 limit set, 458
 limit sets, 40
 Lyapunov exponent, 78, 240, 243, 272, 338, 348, 350, 358, 383, 426, 439–442, 445
 chaos, 338, 392
 gauge invariance, 10, 241, 386, 387, 394
 Kolmogorov entropy, 223, 349
 short time average, 351, 359
 time variable, 332, 350, 359, 365, 433
 transition averaged, 446
 Lyapunov stability, 460
 Lyapunov's Lemma, 34

 Mandelbrot set, 71
 marginally chaotic, 338
 Maupertius Principle, 352, 396–399
 Melnikov method, 87, 149, 206, 211, 213, 354
 stochastic layer, 213

- microwave background radiation, 182, 215, 239, 267
- Milne model, 186
- minisuperspace, 214, 332, 333, 359
 - proper time, 339, 351
- Mixmaster, *see* Bianchi IX, Mixmaster
- non-Abelian gauge fields, 362
- nonlinear pendulum, 213
- numerical techniques
 - Berger-Oliger algorithm, 165
 - convergence tests, 162, 174
 - finite differences, 161
 - Richardson expansion, 163
 - sensitivity, 174
- observer area distance, 244
- oil exploration, 206
- Omega time, 358
- orbit
 - quasiperiodic, 60–62
- Painlevé method, 344, 354, 372, 427–431
- particle creation, 225, 417
- period doubling bifurcation, 131, 139
- periodic orbit
 - hyperbolic, 85
- periodic point, 66, 75
- perturbation
 - dissipative, 207
 - periodic, 210
- phase space compactification, 191
- Planck scale, 223, 336, 425
- Poincaré section, 84, 117, 139, 142, 214, 220, 221, 272, 350
- Poincaré-Bendixson Theorem, 50
- Poincaré metric, 81
- polka dot models, 279
- quantum cosmology, 230–231, 404–413
- quantum maps, 231
- quasi-periodic, 395
- quaternions, 319
- Raychaudhuri equation, 185, 216, 370
- Raychaudhuri Theorem, 185
- resonance condition, 219
- resonance overlap, 214
- scalar field, 183, 215, 304–305, 307
 - collapse, 158
- SDIC, 167
- coupling, 216
- FLRW model, 217
- free, 215
- massive, 208, 214, 240, 304
- massless, 159, 196, 227, 304
 - collapse, 157
 - symmetries, 304
- non-minimal coupling, 176
- oscillations, 170
- spherical collapse
 - critical phenomena, 167
- scalar-tensor theories, 307
- SDIC, 4, 62, 69, 71, 75, 77, 109, 110, 112, 113, 158, 240, 332, 342, 354
 - chaos, 75, 88
 - curvature, 84
 - scalar field collapse, 167, 168
- self similar, 70, 299, 365
- sensitive dependence on initial conditions, *see* SDIC
- separatrix, 14, 149, 207, 210
 - black hole orbit, 212
 - FLRW models, 218
 - Melnikov method, 211
 - Taub, 454–456
- separatrix map, 210–211
- shear, 185
- shock formation, 313
- singularity, 182, 205
 - asymptotically velocity term dominated, 309
 - curvature, 311
 - generic cosmological, 309, 335
 - Gowdy, 309
 - Kasner, 196
 - Mixmaster, 193, 310, 332, 348, 357, 368
 - generic, 400
 - self-similar, 382
 - spatial structure, 310, 313–314, 364
 - theorems, 185, 196, 309
- Smale Horseshoe, 75, 84, 354
- spacetime
 - asymptotically flat, 121
 - self similar, 300
 - topological structure, 253
- stochastic layer, 206, 210, 211, 229
 - black hole orbits, 155, 212

- FLRW model, 217
- Melnikov method, *see* Melnikov method, stochastic layer width, 211
- stochastic sea, 206, 221
- FLRW model, 219
- strange attractor, *see* attractor
- strong causality, 125
- structural stability, 54–57, 117, 240, 400
- superspace, 393
- Swiss-cheese model, 282
- symbolic dynamics, 395
 - Bianchi IX, 390–392
 - correlation length, 396
 - gauge invariance, 395
 - pruning, 391
- symplectic structure, 83, 106, 112, 113
- tangent bifurcation, 139
- Taub model, 215, 349, 373, 382
- Taub separatrix, 454–456
- time
 - and quantum cosmology, 231
 - Bianchi IX, *see* Bianchi IX, time variable
 - invariant meaning, 109
- Lyapunov exponent, *see* Lyapunov exponent, time variable
- time-machine, 115
- Tolman model, 275, 282
 - FLRW as special case, 296
 - fractal solution, 289–292
 - structural stability, 291
 - fractal solutions, 289
 - redshift, 283
- Tolman solution, 186
- topological backscattering, 266
- topological entropy, 78
- topological equivalence, 31, 45
- torus automorphism, 207, 208, 242
- transitive orbit, 69
- transitive point, 75
- trapping set, 42
- variational equations, 77
- vorticity, 184
- Weyl Curvature Hypothesis, 417
- Wheeler-DeWitt equation, 215, 365, 406
- Yang Mills theory, 403
- Yang-Mills theory, 362
- Ziglin's theorem, 106, 431–432