The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.
Topics in Cosmology

Claudia Stacey Hobbs

A Thesis

presented to the University of Cape Town

for the degree of

Master of Science

in

Applied Mathematics

Supervised by

P.K.S. Dunsby

March 2001
Acknowledgements

While researching and writing this thesis, I have received help and support from many more people than I can mention here. In particular, I would like to thank my supervisor Peter Dunsby for his patient help and guidance and for many interesting discussions which both maintained my enthusiasm for and improved my understanding of the topics I have studied.

I would also like to thank the members of the Cosmology research group at UCT for the dynamic research environment and friendly support this group provides. In particular, I would like to thank George Ellis, who introduced me to the ideas which led to the second part of this thesis, and who has helped me with my understanding of the observational topics. Chris Koen and Anthony Fairall have also been involved in the observational research, and their help and experience in this field have been extremely useful.

My thanks are also due to Roy Maartens, Christos Tsagas and Matthias Marklund for helpful discussions concerning the work done on perturbations, and to Ulf Nilsson for kindly allowing me to use some of his dynamical systems pictures.

I would not have been able to devote my energy to writing this thesis if it wasn’t for my friends and family, who are infinitely supportive, and cooked me supper. Thank-you, you know who you are.

Finally, I wish to thank the NRF (South Africa) for financial support.
Abstract

This thesis contains two distinct parts: the first part introduces and explains the relevant theory and background necessary for the analytic work done on magnetized cosmological perturbations at the end of the first part. The second part discusses some issues related to observational cosmology.

After an introductory chapter including an overview of the thesis, PART I starts with a discussion of the covariant approach to cosmology, introducing notation needed in the thesis. The covariant approach to perturbations is then discussed, and the basic inhomogeneity variables describing energy density, pressure and expansion perturbations are introduced. Their exact evolution equations are presented before being linearized about an FRW background.

The basic ideas of dynamical systems analysis of equations are then introduced and illustrated by means of two examples: perturbations in a perfect fluid are analysed in flat and open FRW models, and then the analysis is carried out again, using a 2-fluid approach, also considering both flat and open FRW models. A brief introduction to cosmic magnetic fields motivates their inclusion in an analysis of density perturbations.

Assuming a large-scale homogeneous magnetic field, a covariant and gauge-invariant approach is followed to describe the evolution of density and magnetic field inhomogeneities and curvature perturbations in a matter radiation universe. The governing equations are set up as a fourth order dynamical system using a two-fluid approach. The equilibrium points of the system are analysed in the radiation and dust eras separately. In the dust-dominated era, this study leads to a magnetic critical length scale closely related to the Jeans Length. The solutions show three distinct behaviours depending on the size of the wavelengths relative to this critical scale. PART II ends with a summary of the results.

PART II contains an overview of various aspects of galaxy observations, concentrating on clarifying the various uncertainties present in measurements of the basic observational parameters of apparent luminosity, surface brightness and number counts. The concept of an observational map is introduced and discussed, and an overview of detection limits and various distance dependant selection effects is given. This part concludes with a discussion of a bivariate galaxy luminosity function derived by De Jong and Lacey.
Contents

Acknowledgements ........................................ i
Abstract .................................................. ii
List of tables ........................................... viii
List of figures ........................................... x
Conventions and abbreviations ......................... xi

1 Introduction ............................................. 1
  1.1 Spacetime and Simple Cosmological Models ......... 1
    1.1.1 Friedmann-Robertson-Walker Models .......... 1
  1.2 Observations ..................................... 3
    1.2.1 Mass distribution ............................ 3
    1.2.2 Distance Measurements ...................... 4
  1.3 Structure Formation .............................. 5
    1.3.1 Cosmic Magnetic Fields ..................... 5
  1.4 Outline of Thesis ................................ 6

II ................................................................ 8

2 The Covariant Approach To Cosmology ................ 9
  2.1 Introduction ....................................... 9
  2.2 Observers and Variables .......................... 9
    2.2.1 Observers ..................................... 9
    2.2.2 Kinematical Quantities ..................... 10
  2.3 Matter Description ................................ 12
    2.3.1 Energy momentum tensor .................. 12
    2.3.2 The imperfect fluid case ................. 13
    2.3.3 The perfect fluid case .................... 14
  2.4 Spacetime Geometry ................................ 14
2.4.1 The Einstein Field Equations .................................. 14
2.4.2 The Ricci Identities ........................................... 15
2.4.3 The Bianchi Identities ......................................... 16
2.5 Propagation and Constraint Equations .............................. 16
2.5.1 Energy and Momentum Conservation ............................ 16
2.5.2 Propagation equations .......................................... 17
2.5.3 The Constraint Equations ...................................... 18
2.5.4 Maxwell-like gravitational propagation and constraint equations ........................................... 19
2.5.5 Maxwell field equations ........................................ 20
2.5.6 Irrotational flow: $\omega = 0$ .................................. 21

3 The Covariant Approach to Cosmological Perturbations ............ 23
3.1 Introduction ......................................................... 23
3.2 Gauge Invariance .................................................. 24
3.2.1 Gauge Transformations ....................................... 25
3.2.2 Specification of $\Phi$ .......................................... 26
3.2.3 Stewart and Walker Lemma ................................. 28
3.3 Basic Inhomogeneity Variables ................................... 29
3.4 Exact Non-linear Evolution Equations ............................. 30
3.4.1 Propagation equation for the density gradient ............... 30
3.4.2 Propagation equation for the expansion gradient ........... 31
3.5 Almost FRW Universes .............................................. 31
3.5.1 Properties of FRW models ................................. 32
3.5.2 Covariantly defined GI variables ............................ 33
3.5.3 Linearization .................................................... 34
3.6 Linear Evolution Equations ....................................... 37
3.6.1 Energy and momentum conservation .......................... 37
3.6.2 Linear Hydrodynamic equations ............................. 38
3.6.3 Linear constraint equations .................................. 39
3.6.4 Linear Maxwell-like gravitational equations ............... 40
3.6.5 Linear equations for the spatial gradients .................. 40
3.6.6 Consistency of the linear Equations ......................... 43
3.7 Scalar Variables .................................................... 43
3.7.1 Local decomposition ......................................... 43
3.7.2 Equations for scalar variables .............................. 44
3.8 Local Classification of Perturbations ............................. 46
3.9 Solutions and Harmonic Components ............................ 46
3.9.1 Solutions ....................................................... 47
3.9.2 Radiation ........................................ 48

4 Dynamical Systems Theory ........................................ 49
4.1 General Introduction ........................................ 49
4.2 Analysis ........................................ 50
  4.2.1 Phase spaces and trajectories ......................... 50
  4.2.2 Equilibrium Points ................................... 51
  4.2.3 Invariant sets ..................................... 55
  4.2.4 Non-linear systems .................................. 56

5 Dynamical Systems Analysis of Simple Perturbation Examples ........................................ 57
  5.1 Introduction ........................................ 57
  5.2 Perturbations of FRW Universes ......................... 58
  5.3 FRW Perfect fluid example ............................. 59
    5.3.1 Flat models ..................................... 60
    5.3.2 Open models ..................................... 61
  5.4 FRW 2-fluid approach ................................ 65
    5.4.1 Flat models ..................................... 69
    5.4.2 Open models ..................................... 73
  5.5 Perturbations of Bianchi Universes .................... 74
  5.6 General applications of DS in Cosmology ............... 75

6 Large Scale Magnetic Fields .................................. 77
  6.1 Introduction ........................................ 77
  6.2 Observations ........................................ 78
    6.2.1 Background ...................................... 79
    6.2.2 Current Observations ............................. 80
  6.3 Origins of Cosmic Magnetic Fields ..................... 81
    6.3.1 Possible origins of seed field .................... 82
    6.3.2 Amplification of a seed field .................... 83
  6.4 Magnetic Fields and Structure Formation ............... 84

7 A Dynamical Systems Approach to Magnetized Cosmological Perturbations ......................... 86
  7.1 Overview of Chapter ................................... 86
  7.2 Equations ........................................ 86
    7.2.1 Basic Equations ................................ 86
    7.2.2 Gauge Invariant Variables ....................... 87
7.2.3 Propagation and Conservation Equations ............ 88
7.3 Approximations ........................................... 90
7.4 Linearization .............................................. 91
  7.4.1 Linearization Scheme ................................ 91
  7.4.2 Linear Equations .................................... 93
  7.4.3 Final Linearized System ............................. 95
7.5 The Dynamical System .................................... 96
  7.5.1 Coefficients in terms of $w$ ......................... 97
  7.5.2 Final system ......................................... 99
7.6 Analysis .................................................. 100
  7.6.1 Radiation era: $w = \frac{1}{3}$ ......................... 100
  7.6.2 Dust era: $w = 0$ .................................. 104
  7.6.3 Magnetized Jeans Length ......................... 109

8 Summary of Part I ...................................... 111

II .......................................................... 114

9 Uncertainties in Galactic Observational Parameters 115
  9.1 Observational Elements ................................ 115
  9.2 Measurement Uncertainties: Luminosity ............... 115
    9.2.1 Distance uncertainties .......................... 116
    9.2.2 Uncertainties in the Tully-Fisher Relation .... 118
    9.2.3 Determining absolute magnitudes ............... 121
  9.3 Measurement Uncertainties: Surface Brightness .... 122
    9.3.1 Surface photometry of spiral galaxies .......... 123
    9.3.2 Dust and inclination effects .................... 123
  9.4 Fundamental Plane .................................... 124
    9.4.1 Description .................................... 124
    9.4.2 Determinations of the fundamental plane for spiral galaxies ............................. 125

10 Number Counts and the Galaxy Luminosity Function 128
  10.1 Number Counts ...................................... 128
    10.1.1 Defining the observational map ................. 129
    10.1.2 Detection limits ................................ 131
    10.1.3 Selection effects .............................. 132
  10.2 Galaxy Luminosity Function .......................... 134
10.2.1 Visibility of galaxies ........................................... 136
10.2.2 The two-parameter luminosity function .................. 136
10.3 Summary of Part II ............................................. 138

A Covariant Identities ........................................... 140

B Properties of the orthogonal covariant derivative ............ 141
  B.1 The spatial derivative ....................................... 141
  B.2 The 3-curvature tensor ..................................... 142

C Bianchi and Inhomogeneous Models ............................ 144
  C.1 Isometries and Killing vectors .............................. 144
  C.2 Definitions .................................................. 145
     C.2.1 Bianchi Cosmologies .................................... 145
     C.2.2 $G_2$ Cosmologies ...................................... 145
  C.3 Second order equation for DS analysis of Bianchi I perturbations ........................................ 145
List of Tables

5.1 Equilibrium points for flat ($\Omega = 1$) models, with the corresponding eigenvalues ........................................ 69
5.2 Equilibrium points for open models with the corresponding eigenvalues .......................................................... 73
7.1 Table of equilibrium points for $w = \frac{1}{3}$ ................................................................. 101
7.2 Table of equilibrium points for $w = 0$, with $\xi = \frac{1}{2} \sqrt{\frac{25}{24} - \frac{k_0^2}{k_{2c}^2}}$. 103
List of Figures

3.1 The perturbed density $\delta \mu$ is defined by a mapping $\Phi$ of an idealized world model $\mathcal{S}$ into a more accurate world model $\mathcal{S}'$. $\Phi$ maps surfaces $\bar{\mu} = \text{const}$ from $\mathcal{S}$ into $\mathcal{S}'$, where they can be compared with the actual surfaces $\mu = \text{const}$. ........................................ 25

3.2 This shows the four aspects of the map $\Phi$ ........................................ 27

4.1 Some orbits in the $x - y$ plane for the non-autonomous system described by $\dot{x} = y$, $\dot{y} = x^2 - t$. ........................................ 51

4.2 Some orbits in the $x - y$ plane for the autonomous system described by $\dot{x} = y$, $\dot{y} = x^2$. ........................................ 51

4.3 The behavior of trajectories around a node source with $\lambda_1 \neq \lambda_2$ strictly positive. ........................................ 53

4.4 The behavior of trajectories around a node saddle with $\lambda_1 < 0 < \lambda_2$. ........................................ 53

4.5 The behavior of trajectories around a node sink with $\lambda_1 \neq \lambda_2$ strictly negative ........................................ 53

4.6 The behavior of trajectories around a spiral sink with $\lambda$ imaginary and $\alpha < 0$. ........................................ 53

4.7 The behavior of trajectories around a center with $\lambda$ purely imaginary since $\alpha = 0$. ........................................ 53

4.8 The behavior of trajectories around a spiral source with $\lambda$ imaginary and $\alpha > 0$. ........................................ 53

4.9 The degenerate case, with two linearly independent eigenvectors, and $\lambda_1 = \lambda_2 = \lambda$ strictly positive. ......................... 54

4.10 The degenerate case, showing a node sink with $\lambda_1 = \lambda_2 = \lambda$ strictly negative. ........................................ 54

4.11 The behavior of trajectories around a line of sources, occurring when $\lambda_1 > 0$, $\lambda_2 = 0$. ........................................ 54
4.12 The behavior of trajectories around a node sink
with $\lambda_1 \neq \lambda_2 \neq \lambda_3$ strictly negative. ................. 55

4.13 The behavior of trajectories around a spiral sink
with $\lambda_1 < 0, \alpha < 0$. ........................................ 55

4.14 The behavior of trajectories around a line of saddles
with $\lambda_1 < 0 < \lambda_3, \lambda_2 = 0$. ........................... 55

5.1 Phase space for open models, showing the case with two ε-
equilibrium points for $\Omega = 0$. .................................. 64

5.2 Phase space for open models, showing the case where the
future attractor is the periodic orbit $\Omega = 0$ ...................... 65

5.3 Phase space showing large scale perturbations in flat models.
Here, as in the other figures $0 \leq w \leq \frac{1}{3}$, and $-4 \leq \mathcal{U} \leq 4$ ....... 70

5.4 Phase space for flat models, showing the oscillatory behaviour
of the perturbations on small scales. ................................. 72

5.5 Phase space for the intermediate regime, showing $\mathcal{U}_+ < 0$ .... 72

5.6 Open model phase space; topology is independent of the wave-
length; all trajectories tend to $\mathcal{U}_+ = 0$ .......................... 74

7.1 Radiation dominated era ........................................... 103

7.2 Dust dominated era - $k_F \leq k_{EC}$ ............................... 106

7.3 Dust dominated era - $k_{EC} < k_F \leq \frac{5}{2}\sqrt{6}k_{EC}$ .......... 107

7.4 Dust dominated era - $k_F > \frac{5}{2}\sqrt{6}k_{EC}$ ................... 109

7.5 Dust dominated era - $k_F > \frac{5}{2}\sqrt{6}k_{EC}$ ................... 110

9.1 A crude plot of the Matthewson data, showing a basic trend
toward a planar distribution ........................................... 126
Conventions and abbreviations

Sign conventions

Signature: \([-, +, +, +, +]\).  
Riemann tensor:  
\[ V^a_{;bc} - V^a_{;cb} = R^a_{dcb} V^d, \]
where ; denotes covariant differentiation with respect to the metric tensor.  
Ricci tensor:  
\[ R_{ab} = R^c_{abc}. \]
Units:  
\[ c = 8\pi G = \kappa = 1. \]
Latin and Greek indices assume the values 0, 1, 2, 3 and 1, 2, 3 respectively.  
Sign conventions follow those of Ellis (1971) and Ellis & van Elst (1998) [38, 44].  
For a tensor \( T^{a_1\ldots a_i} \) we have:  
symmetrization:  
\[ T^{a_1\ldots a_i} \]  
anti-symmetrization:  
\[ T^{a_1\ldots a_i} \]

The Newtonian operators of div and curl can be generalized by defining the covariant spatial divergence [96, 69]:

\[
\text{div } V = \tilde{\nabla}^a V_a, \quad (\text{div} T)_a = \tilde{\nabla}^b T_{ab},
\]

and curl:

\[
\text{curl } V_a = \varepsilon_{abc} \tilde{\nabla}^b V^c, \quad \text{curl } T_{ab} = \varepsilon_{cd(a} \tilde{\nabla}^c T_{b)}^d.
\]

The completely anti-symmetric pseudotensor \( \eta^{abcd} \) is defined by

\[
\eta^{0123} = (-g)^{-\frac{1}{2}},
\]

where \( g \) is the determinant of the metric \( g_{ab} \).  
The projected permutation tensor \( \varepsilon_{abc} \) is defined by:

\[
\varepsilon_{abc} = \eta_{abcd} u^d.
\]

Abbreviations

GI  
gauge-invariant.  
GIC  
gauge-invariant covariant.  
FRW  
Friedmann-Robertson-Walker.
Chapter 1

Introduction

1.1 Spacetime and Simple Cosmological Models

Modern cosmology examines the large-scale structure and evolution of the physical universe based on the general theory of relativity [105]. In order to study this, we develop simplified cosmological models describing the universe at a particular averaging scale. We consider the usual four-dimensional description of space-time. To define a specific cosmological model, we specify the space-time geometry using a Lorentzian metric $g$ defined on the manifold $\mathcal{M}$. We also specify a fundamental 4-velocity field $u$. Gravity acts as a curvature in spacetime, determining the paths that objects in free-fall will follow (geodesics). The matter content of the universe can be specified by the stress-energy tensor of each matter component $T_{ab}^{(i)}$. Since the distribution of matter in the universe determines the nature of the gravitational fields present, it directly determines the geometry of that space-time. We assume that this interaction is described by the *Einstein field equations* (EFE):

$$R_{ab} - \frac{1}{2} R g_{ab} = T_{ab}, \quad T_{ab} = \sum_i T_{ab}^{(i)}. \quad (1.1)$$

1.1.1 Friedmann-Robertson-Walker Models

The solutions to the field equations depend on the distribution of mass in the system under consideration. If this system is the entire Universe, the inhomogeneous distribution of light-emitting material (e.g. galaxies, clusters etc) implies that the geometry is non-trivial, assuming that the mass
distribution is similar to the light distribution. On very large scales, the isotropy of the cosmic background radiation suggests that the universe is approximately homogeneous. This implies that solving Einstein’s Field Equations assuming a homogeneous mass distribution will at least allow one to describe the gross properties of the Universe. Assuming isotropic expansion of the Universe is justified by the lack of any observations which seem to pick out preferred directions in the Universe on very large scales. Under these two assumptions the geometry of the model of the universe is greatly simplified, and is described by the Robertson-Walker metric. The simplest solutions to the field equations for this isotropic, homogeneous geometry are known as Friedman cosmological models with the metric given by [77]:

\[ ds^2 = dt^2 - a^2(t) \left( \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \right), \]

(1.2)

where \( r, \theta \) and \( \phi \) are the usual spherical coordinates, \( K = +1, 0, \) or \( -1 \) and \( a(t) \) is a scaling factor.

Friedmann-Robertson-Walker (FRW) models can be fully characterised by specifying the current rate of expansion of the universe, as well as the current mass density \( \mu_0 \). This is usually expressed as a dimensionless ratio of density to critical density \( \mu_c \), defined as the density parameter

\[ \Omega_0 = \frac{\mu_0}{\mu_c} = \frac{8\pi G\mu_0}{3H_0^2}. \]

(1.3)

If the current mass density is greater than the critical density, then \( \Omega > 1 \) and there is enough mass in the universe for its gravitational influence to eventually cause the universe to stop expanding \(^1\), and to recollapse in on itself [17]. This is known as a closed model. If \( \mu_0 < \mu_c \), then \( \Omega < 1 \) and the universe will expand eternally as an open model. In between these two alternatives, if \( \mu_0 = \mu_c \), the density parameter \( \Omega \) is exactly equal to one, and the universe will still expand forever, but its expansion velocity will tend asymptotically towards zero. This is known as the Einstein-de Sitter model, and is said to have a flat geometry. The amount of matter in the universe also affects the curvature of space, which determines the value of \( K \) in equation (1.2). Open, flat, and closed models have \( K = -1, 0 \) and +1 respectively, corresponding to the spatial hypersurfaces having negative, zero and positive curvature.

\(^1\)This calculation assumes that the universe has no pressure, a simplification which is described by a 'dust' model.
The density of the universe can also be quantified in a slightly more abstract way through the use of the cosmic deceleration parameter $q_0$. This specifies, in a dimensionless form, the rate at which the expansion of the universe is currently decelerating. This quantity is clearly closely related to the mass density responsible for the deceleration.

1.2 Observations

In order to be able to meaningfully interpret astronomical observations and observational relations, it is important to understand the relation between source properties and image parameters. The observational map [45] provides a theoretical framework for examining this problem in a systematic way. This is essential for studying the limitations imposed by selection effects and detection limits on observations. The interpretation of observations depends on understanding these effects, essentially by projecting the limits back to the object plane and expressing them in terms of intrinsic source properties. This is only possible if the metric and the nature of any intervening medium is known. In cosmology, it is the metric which we are trying to determine through observations. A kind of recursive procedure can be used to try to solve this problem, whereby successively better estimates of the cosmology lead to better estimates of the observational effects resulting from that cosmology, and thus to a better determination of source properties [45]. The clearest approach analyses the problem in terms of equations which explicitly make the dependences clear. A well-defined observational map does just this.

1.2.1 Mass distribution

The existence of structure in the Universe allows us to probe the distribution of matter on various size scales. The information obtained from these observations is useful when constructing origin scenarios for structure formation in the universe, since the distribution of mass on both large and small scales provides a stringent constraint on any physical theory used to predict the formation and growth of such structure [15]. There are currently serious concerns about how best to measure large scale structure (LSS) from observations. The simplest way to do this is to use galaxies as tracers of the mass. Under the assumption that galaxies are fair tracers of the underlying mass distribution, a direct measure of this distribution can be obtained by
using redshift surveys to determine the three-dimensional galaxy distribution. The positions of the galaxies in redshift space is not necessarily the same as physical space, since mass concentrations distort the appearance of structure when plotted in this way. The evidence for non-luminous matter in clusters of galaxies must also be taken into account. Although the precise nature of this dark matter and its possible distribution is unknown, it definitely affects the relation between redshift space and physical space.

If we assume that the light distribution is a fair and unbiased tracer of the mass distribution, then lightless areas (voids) are also massless areas. The highly clustered nature of the light distribution would then reflect a highly clustered mass distribution. If the distribution of mass is smoother than the distribution of light, then bias would be introduced, assuming that light traces mass. The simplest form of biasing is a linear relation between the galaxy density and the dark matter density [15]:

$$\delta_{\text{galaxies}}(r) = b\delta_{\text{dark matter}}(r),$$  \hspace{1cm} (1.4)

where $b$ is the linear biasing factor. Here, $b$ represents a large-scale mean biasing value. Larger or smaller values of $b$ on smaller scales can exist. The issue of biasing stems from our lack of knowledge about the nature of dark matter and the process of structure formation.

### 1.2.2 Distance Measurements

In order to determine one of the most fundamental cosmological parameters, the current mass density of the universe $\mu_0$, it is necessary to measure volumes of space and hence the distances of any sources which are observed. These distances are not directly observable, and the extragalactic distance scale is built largely on secondary distance indicators, which need to be well calibrated.

One of the most important cosmological results, the Hubble law, was originally obtained without good absolute distance calibration [17]. The law states that the redshift $z$ of a galaxy is proportional to its distance $D$.

$$D = \frac{cz}{H_0},$$  \hspace{1cm} (1.5)

where $c$ is the speed of light. Without good absolute distance measurements, the Hubble constant, $H_0$, is an ill-determined constant of proportionality. It is usually written in the form

$$H_0 = 100h \text{km s}^{-1}\text{Mpc}^{-1},$$  \hspace{1cm} (1.6)
where the dimensionless constant $h$ parameterizes our ignorance of the value of the Hubble constant, and has a true value somewhere between 0.64 and 0.8 [52].

1.3 Structure Formation

The overall structure of the observable universe is assumed to be well described by the FRW models discussed above. This assumption leads to structure formation scenarios in which galaxies, clusters, superclusters and voids evolved somehow from initially small density perturbations in a FRW background [27]. Generic features in the analysis of structure formation make it difficult to use observations to test the models that describe the origins of the inhomogeneity. Regardless of whether these small perturbations were primordial, or spontaneously arose at some very early epoch, they evolved linearly for a long time. Two distinct perturbation modes can be identified. Starting with uniform distributions of matter and radiation, the simplest way to perturb the density is to compress or expand some set of volume elements adiabatically. This changes the matter and photon number densities by the same factor, but changes the energy densities of matter, $\mu_m$, and radiation, $\mu_r$, in different ways, since $\mu_r \propto T^4$ while $\mu_m \propto T^3$. These energy density perturbations are termed adiabatic perturbations. An alternative approach is to perturb the entropy density, but not the energy density. These fluctuations are known as isocurvature perturbations since there is no perturbation to the spatial curvature due to the total density remaining homogeneous [78].

1.3.1 Cosmic Magnetic Fields

Advances in observational techniques have shown that magnetic fields are widespread in the universe. Magnetic fields have been detected in planets, stars, galaxies and clusters of galaxies. Primordial fields may have had effects on structure formation, since a seed field with a very small magnitude, of the order of $10^{-12}$ to $10^{-9} G$ may still produce significant magnetic fields in galaxy clusters through adiabatic compression. The strength of any primordial magnetic field is strongly limited by astrophysical constraints such as nucleosynthesis considerations and the isotropy of the cosmic microwave background, which effectively rule out the possibility of a strong homogeneous large-scale field. The origin of any kind of cosmic seed field is still undetermined, as shown by the existence of at least 30 theories concerning
the origin of cosmic magnetic fields at galactic and intergalactic scales [13]. Even so, magnetic fields can be considered an additional consideration in structure formation scenarios.

1.4 Outline of Thesis

This thesis is presented in two parts: the first part (up to chapter 7) deals with some theoretical aspects of large scale structure formation in a Friedmann-Robertson-Walker (FRW) universe and the second part deals qualitatively with the problems and uncertainties involved in galaxy observations, particularly those that are relevant to the determination of structure formation scenarios and cosmological parameters. PART I starts with a chapter outlining the covariant approach to cosmology as presented in the papers by Ehlers [37], Ellis [38, 44] and Hawking [56]. The chapter introduces the notation used elsewhere in the thesis, particularly the notion of a fundamental observer, and the kinematical quantities derived from the decomposition of the covariant derivative of the 4-velocity. Chapter 3 uses the covariant approach to examine cosmological perturbations and the problems of gauge invariance and gauge transformations are discussed. The basic inhomogeneity variables used in the thesis are presented, as well as a brief derivation of their exact evolution equations for an imperfect fluid, following [27]. These equations are then linearized around an FRW background. A local decomposition of the spatial gradient of the density extracts a scalar variable relevant to structure formation and evolution equations are given for this and related variables.

Chapter 4 introduces the basic ideas of dynamical systems theory, and describes the classification of equilibrium points. These concepts are concretely illustrated in chapter 5, in which a dynamical systems analysis of two simple perturbation examples is presented. The first example illustrates the analysis of an FRW perfect fluid in both a flat and open model, following Dunsby (see [111], chapter 14). The second example follows the analysis, by Bruni and Piotrokovska [8] of a FRW 2-fluid approach to perturbations, also considering both flat and open models.

Chapter 6 discusses large scale magnetic fields, describing current observations and origin scenarios. A short overview of the work done in investigating structure formation in the presence of magnetic fields is presented. All the approaches and concepts of the first 6 chapters are combined in chapter 7, in which a dynamical systems approach to magnetized cosmological
perturbations is described in detail, based on a paper by Hobbs and Dunsby [59]. This covariant approach is used to set up the propagation equations for magnetized density perturbations. These equations were first presented in work done by Tsagas and Barrow [96, 97, 98]. A 2-fluid approach [8] is followed, and the stability of the model is investigated using dynamical systems techniques. PART I ends with a short summary of the work presented.

PART II is a brief introduction to some aspects of galaxy observations. Chapter 9 introduces the observational elements and basic observations of apparent size and surface brightness which quantify the galaxy population. The uncertainties associated with these basic measurements are explained, concentrating on the those relevant in determining the Tully-Fisher relation [101], one of the most used and most reliable extragalactic distance indicators. Sample biasing, which also affects distance determinations, is discussed. The Fundamental Plane [34], an empirical correlation between central velocity dispersion, effective radius and effective surface brightness for early-type galaxies is described. Some determinations of a similar plane for spiral galaxies are presented.

Number counts, the other basic observation for any source type, are discussed in chapter 10. The observational map is introduced as a way to clarify the relationship between fundamental source properties and observed image properties. Selection and detection effects, both of which affect number counts, are discussed. The galaxy luminosity function, an analytical expression specifying the number of galaxies in a certain magnitude range, is introduced. The need for a two-parameter galaxy function is motivated, and De Jong and Lacey's [36] derivation of a bivariate expression for the space density of galaxies in the luminosity-scale-size plane is presented.

This thesis consists mainly of review work. The research presented in chapter 7, however, is original, based on an accompanying paper by myself and P.K.S. Dunsby.
Part I
Chapter 2

The Covariant Approach To Cosmology

2.1 Introduction

In this chapter, we review the covariant approach to cosmology. This approach was first presented in the classic paper by Ehlers [37], as well as by Ellis [38, 44] and Hawking [56]. We will be working in the streamlined notation introduced by Maartens in [69]. The material in this chapter is drawn mostly from [44, 27].

2.2 Observers and Variables

2.2.1 Observers

We can think of two families of worldlines representing the 4-velocities of sets of observers $O_u$ and $O_n$. These worldlines, $u^a$ and $n^a$ respectively, are unit future directed timelike smooth vector fields, such that $u^a u_a = n^a n_a = -1$. At each point of spacetime we can define the following projection tensor into a subspace $H_p$ of the tangent space $T_p$ at $p$:

$$h_{ab} = g_{ab} + u_a u_b, \quad \tilde{h}_{ab} = g_{ab} + n_a n_b.$$  \hspace{5cm} (2.1)

These define the spatial part of the instantaneous rest spaces of the observers $O_u$ and $O_n$. These tensors are metrics in the subspaces $H_p$ of the tangent space $T_p$ orthogonal to the corresponding vector field ($u^a$ or $n^a$). The projection tensor $h_{ab}$ (or $\tilde{h}_{ab}$) is the metric in the surface if the vector
field \( u^a \) (or \( n^a \)) is hypersurface orthogonal.\(^1\) The relation between \( u^a \) and \( n^a \) can be characterized by the hyperbolic tilt angle \( \beta \) [62] such that
\[
u^a n_a = - \cosh \beta, \quad \beta \geq 0. \tag{2.2}
\]
The direction \( \bar{c}^a \) of the motion of \( O_u \) in the instantaneous rest space of \( O_n \) is the projection of \( u^a \) into this space. Similarly, the direction \(-c^a\) of the motion of \( O_n \) is the projection of \( n^a \) into the instantaneous rest space of \( O_u \).
This gives
\[
\begin{align*}
\tilde{h}^{a}_{b} u^b &= \sinh \beta \bar{c}^a = \bar{V}^a \Rightarrow \bar{V}^a n_a = \bar{c}^a n^a = 0, \quad \bar{c}_a \bar{c}^a = 1, \tag{2.3} \\
h^a_{b} n^b &= - \sinh \beta c^a = V^a \Rightarrow V^a n_a = c_a u^a = 0, \quad c_a c^a = 1, \tag{2.4}
\end{align*}
\]
Either of these directions can be used to specify the direction of tilt.

The tilt angle \( \beta \) is related to the relativistic contraction factor \( \gamma \) by
\[
\gamma \equiv \cosh \beta = (1 - v^2)^{-\frac{1}{2}}, \quad v = \tanh \beta \tag{2.5}
\]
so that \( \beta \simeq v \ll 1 \) corresponds to a non-relativistic relative velocity \( v \) between \( O_u \) and \( O_n \). In this case
\[
d^a \equiv u^a - n^a \simeq \beta c^a \equiv \bar{V}^a \simeq \beta c^a \equiv - V^a, \quad h_{ab} \simeq h_{ab} + 2u_a V_b. \tag{2.6}
\]
We can refer to the change between two arbitrary frames \( u^a \) and \( n^a \) with a small relative velocity as a change of first order in \( \beta \).

### 2.2.2 Kinematical Quantities

In cosmology, the average motion of matter at each spacetime event defines a unique 4-velocity vector
\[
u^a = \frac{dx^a}{d\tau}, \quad u^a u_a = 1. \tag{2.7}
\]
This vector field is tangent to the \emph{fundamental fluid-flow lines}, a congruence of worldlines carrying the \emph{fundamental observers}. Here \( \tau \) is the proper time measured along the fundamental worldlines.

The projection into the local rest spaces of comoving observers is given by:
\[
h_{ab} = g_{ab} + u_a u_b
\]
with
\[
h^a_b h^b_a = h^a_a, \quad h^a_a = 3, \quad h_{ab} u^b = 0. \tag{2.8}
\]
\(^1\)This is discussed further in section 2.5.6 which describes the case of irrotational flows.
The covariant derivative of any tensor $T_{ab\,cd}^{\text{cd}}$ may be split into a comoving time derivative,
\[ T_{ab\,cd}^{\text{cd}} = u^{e} \nabla_{e} T_{ab\,cd}^{\text{cd}} \]  
and a fully orthogonally projected covariant derivative [44]
\[ \widetilde{\nabla}_{e} T_{ab\,cd}^{\text{cd}} = h_{e}^{a} h_{b}^{c} h^{p}_{\,d} h^{q}_{\,b} \nabla_{e} \nabla_{r} T_{f\,pq}^{\text{cf}} \]  
with total projection on all free indices. If $u^{a}$ has non-zero vorticity, $\nabla$ is not a proper 3-dimensional covariant derivative. Angle brackets denote orthogonal projection of vectors and the projected symmetric trace-free (PSTF) part of tensors [69]:
\[ V^{(a)} = h_{a}^{b} V^{b}, \quad T^{(ab)} = [h_{c}^{a} h_{b}^{d} - \frac{1}{3} h_{ab} h_{cd}] T^{cd}. \]  
Following [95, 69], we can define a covariant spatial divergence and curl that generalize the Newtonian operators to curved spacetime:
\[ \text{div} \, V = \nabla^{a} V_{a}, \quad (\text{div} \, T)_{a} = \nabla^{b} T_{ab}, \]
\[ \text{curl} \, V_{a} = \epsilon_{abc} \nabla^{b} V^{c}, \quad \text{curl} \, T_{ab} = \epsilon_{cde} (\nabla^{d} T_{eb}). \]
Here,
\[ \epsilon_{abc} \equiv \eta_{abcd} u^{d} \Rightarrow \epsilon_{abc} = \epsilon_{[abc]}, \quad \epsilon_{abc} u^{c} = 0 \]
is the projected covariant permutation tensor of the spacetime. It is a volume element for the rest spaces and satisfies the following identities [69]:
\[ \epsilon_{abc} \epsilon_{efg} = 3! h_{d}^{[a} h_{e}^{b} h_{f}^{c]} \]
\[ \epsilon_{abc} \epsilon_{efg} = 2 h_{f}^{[b} h_{g}^{c]} \]
\[ \epsilon_{abc} \epsilon_{abg} = 2 h_{g}^{c} \]
\[ \epsilon_{abc} \epsilon_{abc} = 3! \]
The projection tensor, together with the velocity 4-vector, split the covariant derivative of $u_{a}$ into irreducible basic kinematic quantities [96] [44]:
\[ \nabla_{a} u_{b} = -u_{a} \dot{u}_{b} + \tilde{\nabla}_{a} u_{b} = -u_{a} \dot{u}_{b} + \frac{1}{3} \Theta h_{ab} + \sigma_{ab} + \omega_{ab}. \]
The vector $\dot{u}_{b} = u^{a} \nabla_{a} u_{b}$, is the acceleration vector; it represents the degree to which matter moves under non-gravitational forces, for example pressure. It vanishes for matter in free fall (i.e. moving only under gravity plus inertia).
The trace $\Theta = \nabla^a u_a$ is called the expansion scalar (volume expansion) and represents the isotropic expansion of the fluid. It can be used to define a representative length scale $a$ along the observers' worldline which in turn (in a FRW universe model) defines the Hubble expansion rate as follows:

$$\frac{\dot{a}}{a} = \frac{1}{3} \theta = H$$

(2.19)

The shear tensor $\sigma_{ab}$ is the trace-free symmetric part of $\nabla_a u_b$:

$$\sigma_{ab} = \nabla_{(a} u_{b)}, \quad \sigma_{ab} = \sigma_{(ab)}, \quad \sigma_{ab} u^b = 0, \quad \sigma^a_{\ a} = 0,$$

(2.20)

describing the rate of distortion of the matter flow. The shear magnitude is

$$\sigma^2 = \frac{1}{2} \sigma_{ab} \sigma^{ab} \geq 0, \quad \sigma = 0 \Rightarrow \sigma_{ab} = 0.$$

(2.21)

The skew-symmetric vorticity tensor $\omega_{ab}$ describes the rotation of matter relative to a non-rotating frame.

$$\omega_{ab} = \nabla_{[a} u_{b]}, \quad \omega_{ab} = \omega_{[ab]}, \quad \omega_{ab} u^b = 0.$$

(2.22)

Its magnitude is:

$$\omega^2 = \frac{1}{2} \omega_{ab} \omega^{ab} \geq 0.$$ 

(2.23)

The vorticity can also be represented by the vorticity vector:

$$\omega^a = \frac{1}{2} \epsilon^{abc} \omega_{bc} = \epsilon_{abc} \omega^c,$$

$$\omega_a u^a = 0,$$

$$\omega \cdot 0 \Rightarrow \omega^a = 0 \Rightarrow \omega_{ab} = 0,$$

(2.24)

2.3 Matter Description

2.3.1 Energy momentum tensor

In relativistic thermodynamics [60, 85], the non-equilibrium state of the fluid is described by the energy momentum tensor (EMT) $T^{ab}$, the particle flux $N^a$, and the entropy flux $S^a$. $T^{ab}$ and $S^a$ respectively satisfy the energy-momentum conservation laws and the second law of thermodynamics:

$$\nabla_b T^{ab} = 0, \quad \nabla_a S^a \geq 0$$

(2.25)
A unique timelike and unit eigenvector $u^a_E; (u^a_E u^b_E = -1)$ can be defined from the assumption that the energy density is non-negative i.e.

$$T_{ab} V^a V^b \geq 0$$  \hspace{1cm} (2.26)

for all timelike $V^a$. Another unit timelike vector $u^a_N$ can be defined from thermodynamic variables as $u^a_N = N^a/\sqrt{-g^a N^b N^b}$. This is the unit vector parallel to $N^a$.

### 2.3.2 The imperfect fluid case

The choice of hydrodynamical 4-velocity $u^a$ is not necessarily unique. The EMT can be decomposed relative to the chosen $u^a$ in the form

$$T_{ab} = \mu u_a u_b + p h_{ab} + 2 q_{(a} u_{b)} + \pi_{ab}, \hspace{1cm} (2.27)$$

where

$$q_a = -h_{a}^{b} T_{bc} u^{c}, \hspace{0.5cm} \pi_{ab} = h_{a}^{c} h_{b}^{d} T_{cd} - \frac{1}{3} h_{ab} (h_{cd} T^{cd})$$  \hspace{1cm} (2.28)

are the energy flux, or relativistic momentum density, $q^a u_a = 0$, and anisotropic pressure (stresses) $\pi_{ab} u^a = 0$ in the frame of $u^a$. The isotropic pressure and relativistic energy density are, respectively,

$$p = \frac{1}{3} h^{a} b T^{b}_{a},$$

$$\mu = T_{ab} u^{a} u^{b}. \hspace{1cm} (2.29)$$

The particle flux $N^a$ can also be decomposed with respect to $u^a$ giving

$$N^a = n u^a + j^a, \hspace{1cm} (2.30)$$

where $j^a = h^a_{b} N^{b}$ ($j^a u_a = 0$) is the particle drift in this frame. Two frames which have a special status are known as the energy frame, $u^a_E$, and the particle frame, $u^a_N$. In the energy frame $u^a_E$ there is no energy flux $q^a_E = -h^a_{b} T_{bc} u^c_E = 0$ while in the particle frame the particle drift vanishes $j^a_N = h^a_{N b} N^b = 0$. In general, for an arbitrary non-equilibrium state, there is no simple relationship between the primary thermodynamic variables $S^a$, $N^a$ and $T^{ab}$.
2.3.3 The perfect fluid case

The physics of the fluid is determined by the equations of state relating the quantities $\mu, q^a, \pi^{ab}$, and $p$. The commonly imposed restrictions

$$q^a = \pi_{ab} = 0 \iff T_{ab} = \mu u^a u_b + ph_{ab}$$

(2.31)

characterize a "perfect fluid". This corresponds to a fluid in equilibrium, where $S^a$, $u^a$, $u^b$ are all parallel, and define a unique hydrodynamical 4-velocity for the fluid flow, together with a projector tensor $h_{ab}$ into the local rest space of the fluid. In this case, the decomposition of the three primary thermodynamic variables with respect to $u^a$ has a special status, $u^a$ being the only timelike vector for which $T_{ab}$ takes the perfect fluid form shown in equation (2.31). The particle and entropy flux become

$$N^a = nu^a, \quad S^a = su^a$$

(2.32)

where

$$n = -N_a u^a, \quad s = -S_a u^a$$

(2.33)

are the particle and entropy densities respectively.

The energy density and pressure in the local rest space of $\mathcal{O}$ are

$$\mu = T_{ab} u^a u^b, \quad p = \frac{1}{3} h_{ab} T^{ab}$$

(2.34)

and are related by an equation of state of the form:

$$p = p(\mu, s).$$

(2.35)

It is usually assumed that (2.31) holds in standard FRW spacetimes even if the fluid is not barotropic, so that the matter 4-velocity is uniquely defined in these universe models.

2.4 Spacetime Geometry

2.4.1 The Einstein Field Equations

The metric $g_{ab}$ satisfies the Einstein field equations (EFE):

$$R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} = T_{ab},$$

(2.36)

where $R_{ab} = R^c_{acb}$ is the Ricci tensor obtained by contracting the Riemann tensor (2.37); $R = R^a_a$ is the Ricci scalar obtained by contracting the Ricci
tensor; $T_{ab}$ is the energy momentum tensor of all the matter present and $\Lambda$ is the cosmological constant. These equations determine the trace part of the gravitational field at each point in spacetime from the matter at that point. The trace free part of the gravitational field at each point is determined by the "electric" and "magnetic" parts of the Weyl conformal curvature tensor.

### 2.4.2 The Ricci Identities

The Riemann tensor $R_{abcd}$ represents the curvature of spacetime. It is defined by the Ricci identities, which demonstrate the non-commutativity of second covariant derivatives. For the 4-velocity vector $u^a$ we have

$$2\nabla_a \nabla_b u^c = R_{dab}^c u^d. \quad (2.37)$$

The Riemann tensor has the following symmetry properties:

$$R_{abcd} = R_{[ab][cd]} = R_{cdab}, \quad R_{a[bc]} = 0, \quad (2.38)$$

which give 20 independent components. The Riemann tensor can be algebraically separated into its trace and trace free parts, the Ricci tensor $R_{ab}$ and Weyl tensor $C_{abcd}$ respectively. Both have 10 independent components. The Ricci tensor is determined locally at each point by the matter tensor, through the EFE (2.36), and the Weyl tensor is defined by:

$$C_{abcd} = R_{abcd} - 2g_{[a|c}^{[a} R_{b]d]} + \frac{R}{3} g_{[a|c}^{[a} g_{b]d]}, \quad (2.39)$$

As implied by this definition, the Weyl tensor satisfies all of the symmetries of the Riemann tensor (2.38) and is trace free:

$$C^a_{b\,cd} = 0. \quad (2.40)$$

The Weyl tensor can be split relative to $u^a$ into an 'electric' and a 'magnetic' part. These are defined in analogy with the way the electromagnetic field tensor $F_{ab}$ can be split into electric and magnetic parts. In the case of the Weyl tensor we have:

$$E_{ab} = C_{abcd} u^c u^d \Rightarrow E^a_{\,\,a} = 0, \quad E_{ab} = E_{(ab)}, \quad E_{ab} u^b = 0,$$

$$H_{ab} = \frac{1}{2} \varepsilon_{cde} C_{bde} u^c \Rightarrow H_{a\,\,a} = 0, \quad H_{ab} = H_{(ab)}, \quad H_{ab} u^b = 0. \quad (2.41)$$

These determine gravitational action at a distance (e.g. tidal forces, gravitational waves). They completely represent the Weyl tensor, which we can
write as follows:

\[ C_{abcd} = (\varepsilon_{abq} \varepsilon_{cds} + g_{abpq} g_{cdrs} u^{p} u^{r}) E^{qs} + (\varepsilon_{abq} g_{cdrs} u^{r} + g_{abpq} \varepsilon_{cdst} u^{p}) H^{qs}, \]

(2.42)

where

\[ g_{abcd} = g_{acbd} - g_{adbc}. \]

(2.43)

### 2.4.3 The Bianchi Identities

The Riemann tensor also satisfies the **Bianchi identities**:

\[ \nabla_{[a} R_{bcde]} = 0 \Rightarrow \nabla_{a} R_{bcde} + \nabla_{c} R_{abde} + \nabla_{b} R_{cade} = 0. \]

(2.44)

These can be contracted once to give

\[ \nabla^{a} R_{abcd} = \nabla_{c} R_{bd} - \nabla_{d} R_{bc}. \]

(2.45)

Contracting again gives the **twice contracted Bianchi identities**

\[ \nabla^{a} R_{ac} = \frac{1}{2} \nabla_{c} R. \]

(2.46)

The first of the Bianchi identities (2.44) is equivalent to the field equations (2.36) as there are the same no of components in each set. The twice contracted Bianchi identity shows that the Einstein field equations are only consistent if the conservation equations (2.25) are satisfied by the combined matter fields.

### 2.5 Propagation and Constraint Equations

#### 2.5.1 Energy and Momentum Conservation

These equations can be obtained by projecting either the conservation equation (2.25) or the twice contracted Bianchi identities (2.46) parallel and perpendicular to \( u^{a} \). This splits the equations into a time and a space part, to give

\[ \dot{\mu} + (\mu + p) \Theta + \pi^{ab} \sigma_{ab} + (\text{div } q) + 2 u_{a} q^{a} = 0 \]

(2.47)

and

\[ (\mu + p) \dot{u}^{a} + q^{(a)} + \nabla^{a} p + (\text{div } \pi)^{a} + \dot{u}_{b} \pi^{ab} + \frac{2}{3} \Theta q^{a} + \sigma_{ab} q^{b} + \varepsilon^{abc} \omega_{bc} q_{c} = 0, \]

(2.48)

respectively.
2.5.2 Propagation equations

The evolution equations for the kinematical quantities defined in section 2.2.2 arise from the Ricci identity (2.37). They are obtained by looking at the part along \( u^a \) of the result of substituting in from (2.18), using the EFE (2.36) and separating out the result into a trace, symmetric trace-free and skew symmetric parts.

The Raychaudhuri equation

The evolution equation for the expansion is

\[
\dot{\Theta} - \Lambda = 2(\sigma^2 - \omega^2) + \frac{1}{3}\Theta^2 + \frac{1}{2}(\mu + 3p) - \Lambda = 0,
\]

where \( \Lambda = \nabla^a \dot{u}_a \). Using the definition of the expansion as the relative change in scale factor (2.19), the equation above can be rewritten in the form

\[
\frac{3\dot{\sigma}}{a} = 2(\omega^2 - \sigma^2) + A - \frac{1}{2}(\mu + 3p) + \Lambda.
\]

This shows how the scale factor \( a \) is directly determined at each spacetime point by the matter density at that point, and leads to the identification of \( \mu + 3p \) as the active gravitational mass. It also reveals the repulsive effect of a positive cosmological constant \( \Lambda \), the tendency of vorticity to hold matter apart and contribution of the shear to contraction. The acceleration, which represents spatial pressure gradients, affects the average distance of world-lines through its divergence.

The shear propagation equation

This is the symmetric part of the equation obtained by projecting on the indices \( c \) and \( a \) of the Ricci identity (2.37):

\[
\ddot{\sigma}^{(ab)} - \nabla^{(a} \dot{\sigma}^{b)} + \frac{2}{3} \Theta \sigma^{ab} - \dot{\omega}^{(a} \dot{\omega}^{b)} + \sigma^{(a} \sigma^{bc)} + \omega^{(a} \omega^{b)} + (E^{ab} - \frac{1}{2} \pi^{ab}) = 0,
\]

showing how the anisotropic pressure \( \pi^{ab} \) and the tidal gravitational field \( E^{ab} \) directly induce a shear distortion in the fluid flow.
The vorticity propagation equation

This is derived by taking the antisymmetric part of the equation obtained from the Ricci identity (2.37) by projecting on the indices $c$ and $a$.

$$\dot{\omega}^{(a)} - \frac{1}{2} (\text{curl } \dot{u})^a + \frac{2}{3} \Theta \omega^a - \sigma^a \omega^b = 0. \quad (2.52)$$

2.5.3 The Constraint Equations

These, like the propagation equations, are derived from the Ricci identity (2.37). The constraint equations are those components of the identity which are perpendicular to $u^a$ on the index $b$, and do not involve time derivatives or kinematic quantities.

Shear constraint

This equation is obtained by contracting on the indices $c$ and $a$ of the Ricci identity (2.37):

$$(\text{div } \sigma)^a - \frac{2}{3} \vec{\nabla}^a \Theta + (\text{curl } \omega)^a + 2 \varepsilon^{abc} \dot{u}_b \omega_c + q^a = 0. \quad (2.53)$$

This equation shows how the energy flux vector $q^a$ relates to the spatial inhomogeneity of the expansion, as well as to the spatial gradients of vorticity and shear.

Vorticity constraint

From (2.37) it follows that the identity $R_{a[bcd]} u^a = 0$ implies $\vec{\nabla}_{[c} \vec{\nabla}_{d} u_{b]} = 0$. Multiplying by $\varepsilon^{bcd}$, we can write this identity in the form:

$$(\text{div } \omega) - \omega^a \dot{u}_a = 0. \quad (2.54)$$

H constraint

The third constraint equation is also obtained from the Ricci identity. In this case the equation (2.37) is multiplied by $\varepsilon^{c/d}$ and symmetrizing on $a$ and $d$:

$$H_{ab} + 2 \dot{u}^{(a} \omega^{b)} + \vec{\nabla}^{(a} \omega^{b)} - \text{curl } \sigma^{(ab)} = 0. \quad (2.55)$$

This shows that the magnetic part of the Weyl tensor is related to the curl of the shear and to the velocity distortion.
2.5.4 Maxwell-like gravitational propagation and constraint equations

We have already obtained equations for the kinematic quantities, the energy density and the pressure. Since the change in $\sigma_{ab}$ is determined by the tidal effect of $E_{ab}$ (as in equation (2.51)), we need to specify the evolution of $E_{ab}$ and $H_{ab}$.

Whereas the previous set of constraint and propagation equations were derived from the Ricci identity (2.37), the evolution equations for $E_{ab}$ and $H_{ab}$ come from the once-contracted Bianchi identities (2.45). Using the definition of the Weyl tensor (2.39), the Bianchi identities become:

$$\nabla_d C^{abcd} = \nabla^{[b} R^a]c - \frac{1}{3} g^{[a} \nabla^{b]} R.$$  (2.56)

Substituting equation (2.42) and the EFE (2.36) into the above, we obtain 2 propagation and 2 constraint equations for $E_{ab}$ and $H_{ab}$. The constraint equations involve the divergences of both these quantities.

Divergence of $E$ equation

$$(\text{div } E)^a + \frac{1}{2} (\text{div } \pi)^a - \frac{1}{3} \nabla^a \mu + \frac{1}{3} \Theta q^a - \frac{1}{2} \sigma^a q^b - 3 \omega_b H^{ab}$$

$$- \varepsilon^{abc} \left[ \sigma_{bd} H^d_c - \frac{3}{2} \omega_b q_c \right] = 0.$$  (2.57)

This shows that the source for the divergence in $E^{ab}$ is the spatial gradient of the energy density.

Divergence of $H$ equation

$$(\text{div } H)^a + (\mu + p) \omega^a + 3 \omega_b (E^{ab} - \frac{1}{2} \pi^{ab})$$

$$+ \varepsilon^{abc} \left[ \frac{1}{2} \nabla_b q_c + \sigma_{bd} (E^{bd} + \frac{1}{2} \pi^{bd}) \right] = 0.$$  (2.58)

is the (div $H$)-equation, with source the fluid vorticity.
\[ \dot{E} \text{ and } \dot{H} \text{ equations} \]

The propagation equation for \( E^{ab} \) is:

\[
\begin{align*}
\left( \dot{E}^{(ab)} + \frac{1}{2} \pi^{(ab)} \right) & - (\text{curl } H)^{ab} + \frac{1}{2} \nabla^{(a} q^{b)} = \\
& - \frac{1}{2} (\mu + p) \sigma^{ab} - \Theta \left( E^{ab} + \frac{1}{2} \pi^{ab} \right) \\
& + 3\sigma^{(a}_c \left( E^{b)c} - \frac{1}{2} \pi^{c} \right) - \dot{u}^{(a} q^{b)} \\
& + \varepsilon^{cd(a} \left[ 2 \dot{u}_c H^{b)}_d + \omega_c \left( E^{b)}_d + \frac{1}{2} \pi^{b)}_d \right) \right], \quad (2.59)
\end{align*}
\]

while for \( H^{ab} \) it is:

\[
\begin{align*}
\dot{H}^{(ab)} & + (\text{curl } E)^{ab} - \frac{1}{2} (\text{curl } \pi)^{ab} = -\Theta H^{ab} + 3\sigma^{(a}_c H^{b)c} + \frac{1}{2} \omega^{(a} q^{b)} \\
& - \varepsilon^{cd(a} \left[ 2 \dot{u}_c E^{b)}_d - \frac{1}{2} \sigma^{b)_c q_d - \omega_c H^{b)}_d \right]. \\
& \quad (2.60)
\end{align*}
\]

These equations show how the propagation of the gravitational field is governed by the matter distribution. Taking the time derivative of equation (2.59) gives a term of the form (curl \( H \)): commuting the space and time derivatives and substituting from the \( \dot{H} \)-equation eliminates \( H \), resulting in a term in \( \dot{E} \) and one of the form (curl curl \( E \)), which together give the wave operator acting on \( E \). Similar manipulations on \( \dot{H} \) equation give a wave equation for \( H \). These equations are very similar to the Maxwell equations for the electromagnetic field and for this reason \( E_{ab} \) and \( H_{ab} \) are called the "electric" and "magnetic" parts of the Weyl tensor.

### 2.5.5 Maxwell field equations

The electromagnetic field is represented by the Maxwell field strength tensor \( F_{ab} = F_{[ab]} \). This field is split relative to \( u^a \) into electric and magnetic field parts by the relations [44]:

\[
\begin{align*}
E_a &= F_{ab} u^b \Rightarrow E_a u^a = 0, \quad (2.61) \\
H_a &= \frac{1}{2} \varepsilon_{abc} F^{bc} \Rightarrow H_a u^a = 0. \quad (2.62)
\end{align*}
\]

The Maxwell field equations are

\[
\nabla_b F^{ab} = J^a, \quad \nabla_{[a} F_{bc]} = 0, \quad (2.63)
\]

20
where \( J^a \) is the 4-current which generates the electromagnetic field, can also be split relative to \( u^a \) to give both propagation and constraint equations. The electric and magnetic propagation equations are as follows [44]:

\[
\dot{E}^{(a)} + j^a = \sigma^{ab}E_b + \epsilon^{abc}\omega_cE_b - \frac{2}{3}\Theta E^a \\
+ \epsilon^{abc}u_bB_c + (\text{curl } B)^a,
\]

(2.64)

and

\[
\dot{B}^{(a)} = \sigma^{ab}B_b - \epsilon^{abc}\omega_B B_c - \frac{2}{3}\Theta B^a \\
- \epsilon^{abc}u_B E_c - (\text{curl } E)^a.
\]

(2.65)

The two constraint equations are:

\[
\text{(div } E) - 2\omega^aB_a - q = 0,
\]

(2.66)

and

\[
\text{(div } B) + 2\omega^aE_a = 0.
\]

(2.67)

The 4-current \( J^a \) is contained in equations (2.66) and (2.64) through the charge density \( q = -J^a u_a \) and the projected current \( j^a = J^{(a)} \).

2.5.6 Irrotational flow: \( \omega = 0 \)

When \( \omega = 0 \), the 4-velocity vector \( u^a \) is proportional to a gradient i.e.

\[ u^a = -t^a, \text{ for some local functions } f(x^a), t(x^a). \]

(2.68)

Since \( t, a \) is a vector normal to the surfaces \( t = \text{constant} \), this condition states that \( u^a \) is orthogonal to these surfaces. Thus, the instantaneous rest spaces orthogonal to \( u^a \) mesh together to form a spacelike hypersurface \( t = \text{constant} \) orthogonal to \( u^a \). The metric of these 3-spaces is \( h_{ab} \). The function \( t \) can be thought of as a cosmological time coordinate defined by the fluid flow. If \( \dot{u}^a = 0 \) it can be normalized to measure proper time along the matter world lines. We can define an intrinsic curvature for these 3-spaces using the Ricci identity (2.37) in the surfaces. Thus, for any vector field \( X^a \) in the 3-spaces \( X^a u_a = 0 \), we have:

\[
2\nabla_{[e}\nabla_{d]}X_a = \left. R_{abcd}X^b \right|^{(3)}.
\]

(2.69)
Using this identity in the hypersurfaces, and using (2.36,2.27, 2.51), we obtain the Ricci tensor \( {}^{(3)}R_{ab} \equiv h^{cd}{}^{(3)}R_{abcd} \) of the 3-spaces:

\[
{}^{(3)}R_{ab} = -\dot{\sigma}_{(ab)} - \Theta \sigma_{ab} + \tilde{\nabla}_{(a} \dot{\gamma}_{b)} + \dot{\gamma}_{(a} \dot{\gamma}_{b)} + \pi_{ab} + \frac{1}{3} h_{ab} [2\mu - \frac{3}{2} \Theta^2 + 2\sigma^2 + 2\Lambda].
\]  

(2.70)

We can contract again \( {}^{(3)}R = h^{ab}{}^{(3)}R_{ab} \) to obtain the Ricci scalar of the 3-spaces:

\[
{}^{(3)}R = 2\mu - \frac{3}{2} \Theta^2 + 2\sigma^2 + 2\Lambda,
\]

(2.71)

which is a generalized Friedmann equation, showing how the matter content of the spacetime determines the average curvature of the orthogonal 3-spaces.

Finally, it is worth mentioning that all the dynamic equations presented in this chapter are exact, and can therefore be applied to any fluid flow.
Chapter 3

The Covariant Approach to Cosmological Perturbations

3.1 Introduction

Cosmological perturbation theory plays an important role in attempts to understand the formation of large scale structures in the universe. A number of approaches to the problem of general relativistic perturbations have been presented since the pioneering work of Lifshitz [68]. In 1966, Hawking [56] developed a fully covariant formalism based on the work of Ehlers [37], which dealt with perturbations of the curvature rather than the metric tensor. These works suffered from gauge ambiguities (see section 3.2). A major breakthrough in the study of linear perturbations came with the gauge-invariant (GI) formalism developed by Bardeen [6]. However, many of his key variables do not have a transparent geometrical meaning, the variables are non-local, and the approach is limited to linear perturbations of FRW models [111].

Gauge-invariant quantities must satisfy the Stewart & Walker Lemma [86]. Ellis and Bruni [40] make use of this lemma to define GI variables covariantly in a simple and physically transparent way. They applied their approach to almost FRW universes with pressure-free matter. The approach was extended to treat density perturbations in a perfect fluid universe in two other papers [42, 41]. The gauge-invariant covariant (GIC) approach has also been used to study cosmic background radiation anisotropies (CBR). A GIC version of the Sachs-Wolfe paper [84], based on photon path integration and calculation of the redshift along these paths, is given by Dunsby [29] and
Challinor and Lasenby [25].

The GIC formalism has also been used by Tsagas and Barrow [96, 97, 98] in the development of a GIC approach to the analysis of magnetized density perturbations. This is further investigated in chapters 6 and 7. Bruni and Sonego [10] discuss the issue of observables and gauge invariance as they relate to non-linear spacetime perturbations. The relationship between the GIC approach to perturbations and the Bardeen gauge-invariant formalism is examined in [7].

The geometrical approach to perturbations of Ellis and Bruni [40] has three main advantages over more standard approaches [7]:
(a) it provides a unified treatment for the exact and the linearized theory;
(b) the same GI variables can be used in perturbing different universe models (see section 3.2 for details);
(c) if covariantly defined variables vanish in the background FRW universe, they are also gauge-invariant under large gauge transformations.

This chapter describes the covariant approach to perturbations as applied to an almost FRW universe.

### 3.2 Gauge Invariance

To define perturbations it is necessary to choose a "smooth" one-to-one map between any given background space-time \( \bar{S} \), and the real inhomogeneous, or "lumpy", universe \( S \). The perturbation at any point \( P \) of any quantity is defined as the difference between the actual value of the quantity and its value in the background spacetime at \( P \). For example, an energy density perturbation \( \delta \mu \) is defined as

\[
\delta \mu \equiv \mu - \bar{\mu},
\]

where the only restriction relating the two models is that the perturbation \( \delta \mu \) is small in some suitable sense. If we consider the lumpy universe model \( S \), without knowing how \( \bar{S} \) was used to make the construction, it is impossible to uniquely recover \( \bar{S} \) from \( S \). The definition of the background model in the \( S \) is equivalent to defining a map \( \Phi \) from \( \bar{S} \) to \( S \), taking the density in the smooth model \( \bar{S} \) into a background density \( \bar{\mu} \) \(^1\) in \( S \) [40]. The perturbations defined are thus completely dependent on how that map is chosen.

\(^1\)we use the same symbol for quantities in \( S \) and their images in \( S \), e.g. denoting the image \( \Phi(\bar{\mu}) \) in \( S \) of \( \bar{\mu} \) in \( \bar{S} \) simply by \( \bar{\mu} \).
freedom of choice of this mapping between the background spacetime \( \tilde{S} \) and the perturbed universe \( S \) is referred to as the gauge freedom in defining the perturbation. A change in this mapping is called a gauge transformation.

The situation is usually expressed in terms of the coordinate choice in \( S \) [40]. It is understood that the coordinates in \( S \) correspond to coordinates chosen in \( \tilde{S} \), so that the choice of coordinates determines a map from \( \tilde{S} \) into \( S \). The gauge freedom is thus represented as a freedom of coordinate choice in \( S \). It is clearer to consider the map \( \Phi \) from \( \tilde{S} \) into \( S \) specifically. Coordinate freedom in both \( \tilde{S} \) and \( S \) can then be exploited to adapt to the chosen map \( \Phi \).

### 3.2.1 Gauge Transformations

When the gauge freedom is described in terms of coordinates, gauge transformations are represented by infinitesimal coordinate transformations such that [27]

\[
\tilde{x}^a \rightarrow \tilde{x}'^a = \tilde{x}^a - \epsilon^a(x),
\]

where \( \epsilon^a(x) \) is an arbitrary infinitesimal vector field. A coordinate independent description of the gauge problem is also possible [57]. The above transformation can be interpreted in two ways. It can be thought of as a
relabeling of the point $x$. The change induced by this on scalars, vectors and tensors can then be computed using the usual transformation rules, and neglecting terms of $\mathcal{O}(2)$ or higher. Alternatively, to compute the effect of (3.2) on any scalar, vector or tensor $\tilde{T}$ we can expand it about the point $x$, and then compare the result at the same coordinate point $x$. In either case, the important point is that a gauge transformation changes the point in the background spacetime which corresponds to a specific point in the physical spacetime. The gauge transformation (3.2) thus induces a change in any tensor $\tilde{T}$ such that

$$\tilde{T}'(x) = \tilde{T}(x) + L_\epsilon \tilde{T}(x)$$

(3.3)

where $L_\epsilon \tilde{T}$ is the Lie derivative of $\tilde{T}$ along $\epsilon$. For scalars, vectors and (second-rank) tensors, we obtain [27]:

$$L_\epsilon f = f_a \epsilon^a,$$

(3.4)

$$L_\epsilon V_a = V^b \epsilon_b, a + V_{ab} \epsilon^b,$$

(3.5)

and

$$L_\epsilon T_{ab} = T_{abc} \epsilon^c + T_{ade} \epsilon^e, b + T_{eb} \epsilon^e, a.$$  

(3.7)

From these equations we see that the transformation (3.2) is an infinitesimal diffeomorphism, so that the solutions $\tilde{g}_{ab}$, $\tilde{f}_{ab}$ with sources $\tilde{T}_{ab}$, $T'_{ab}$ are physically equivalent, but have different values at the given coordinate point. Consequently, the value of the perturbation in each quantity is similarly changed under the same transformation. This is the source of the gauge problem, which has been outlined here within a coordinate based approach. A particular gauge choice is equivalent to choosing a map $\Phi$ between the background spacetime $\tilde{S}$ and the physical spacetime $S$. We will examine the properties of this map more closely.

### 3.2.2 Specification of $\Phi$

The actual situation is that the real "lumpy" universe $S$ is available for study, and is, of course, all we can measure. The perturbed quantities and their evolution are then defined by the way the mapping $\Phi$ is specified. Determining the best way to make this correspondence, which is not unique, is called the "fitting problem" in cosmology. The mapping can be thought of as having four aspects [40, 27]:

---

26
• (A) We define a family of world lines $\tilde{\gamma}$ in $\tilde{S}$ and a corresponding family of world lines $\gamma$ in $S$. This determines the worldlines in each spacetime along which we will compare the evolution of fluctuations (e.g., density perturbations). The obvious choice in $\tilde{S}$ is the fundamental flow lines, which is often also the best choice in $S$. Other choices, such as the normals to a chosen set of surfaces, may sometimes be convenient.

• (B) We define a specific correspondence between individual world lines $\tilde{\gamma}_i$ in $\tilde{S}$ and $\gamma_i$ in $S$. This defines which specific observer's observations we shall compare with which. When $\tilde{S}$ is a FRW or a Bianchi model, the spatial homogeneity of these model makes this choice irrelevant.

• (C) We define a family of spacelike surfaces $\tilde{\Sigma}$ in $\tilde{S}$; these are "time surfaces" in each spacetime. Again, there is an obvious choice in $\tilde{S}$, namely the surfaces of homogeneity ($\tilde{t} = \text{const}$). The image of these surfaces in $S$ (i.e., the surfaces $\tilde{t} = \text{const}$ in $S$) are the idealized surfaces of $t$ constant density ($\tilde{\mu} = \text{const}$) which we use to define density perturbations.

• (D) We define a correspondence between particular surfaces $\tilde{\Sigma}_i$ in the family $\tilde{\Sigma}$ in $\tilde{S}$ and particular surfaces $\Sigma_i$ in the family $\Sigma$ in $S$, and so assign particular time values $\tilde{t}$ to each event $q$ in $S$. This crucial choice specifies which specific point $q$ in $S$ corresponds to a point $\tilde{q}$ in $\tilde{S}$, and completes the specification of the map $\Phi$. In particular, this
defines the time evolution of the density perturbation \( \delta \mu \) by assigning particular values \( \bar{\mu} \) to each surface \( \Sigma_i \) in \( S \), and defines \( \delta \mu \) via equation (3.1).

Following the normal convention, (C) is understood to define the coordinate surfaces \( t = \text{const} \) in \( S \), taking them to be the same as the surface \( \bar{t} = \text{const} \); (D) is understood to assign particular values to \( t \) at each event \( q \) in \( S \) by the map \( t_q = \bar{t}_q \). This choice is not, however, forced on us. In general neither \( t \) nor \( \bar{t} \) will measure proper time along the world lines in \( S \).

The freedom in choosing the mapping \( \Phi \), particularly the freedom to choose the surface \( \bar{\Sigma} \) in \( S \), and to allocate density values to this surface, means that the definition of \( \delta \mu \) depends crucially on these choices. In fact, given a choice of the family of surfaces \( \bar{\Sigma} \) in \( S \), any desired value can be allocated to \( \delta \mu \) at a particular event through the gauge freedom (D). This is done by changing the allocation of the values \( \bar{\mu} \) to the surfaces \( \bar{\Sigma} \). This arbitrariness in the definition of \( \delta \mu \) shows that \( \delta \mu \) is not gauge invariant [40, 27]. Not only can \( \delta \mu \) be assigned any value we like at any event through an appropriate gauge choice, but it is also not observable even in principle, unless the gauge is fully specified by an observationally based procedure (without which \( \bar{\mu} \) is not an observable quantity).

Hence, if \( \delta \mu \) is to be used in a satisfactory way to describe density perturbations, either some gauge freedom must be left, and full track kept of the consequences of all this freedom, or a satisfactory, unique way must be found of making the gauge choices (A) to (D) described above. The alternative, described in this chapter, is to use gauge-invariant quantities which code the information wanted. The approach relies on the gauge invariance lemma of Stewart and Walker [86].

3.2.3 Stewart and Walker Lemma

This states [86]:

**Lemma:** The linear perturbation \( \delta \mathcal{T} \) of a quantity \( \mathcal{T}_0 \) on the background spacetime \( \bar{S} \equiv \{ \mathcal{M}_0, g_0 \} \) is gauge invariant (GI) if and only if one of the following holds:

\begin{itemize}
  \item[a)] \( \mathcal{T}_0 \) vanishes;
  \item[b)] \( \mathcal{T}_0 \) is a constant scalar;
  \item[c)] \( \mathcal{T}_0 \) is a constant linear combination of Kronecker deltas.
\end{itemize}
These conditions are those for which the Lie derivative of $\mathcal{T}_0$, $\mathcal{L}_X \mathcal{T}_0$, vanishes for all vector fields $X$ on $\mathcal{M}$. When gauge transformations are described in a co-ordinate independent way (see [27]), it is clear that the above derivative must vanish in order to give gauge invariance. If we consider only the first condition for a background quantity $\mathcal{T}$, we can see that if $\delta \mathcal{T} = 0$, then

$$\delta \mathcal{T} = \mathcal{T} - \mathcal{T} = \mathcal{T},$$

which is independent of the mapping $\Phi$ from $\mathcal{S}$ to $\mathcal{S}$. In the context of this thesis, the only useful quantities are those that vanish in the background spacetime, and we will use these to define a set of covariantly defined GI variables.

### 3.3 Basic Inhomogeneity Variables

Generally, for any scalar $f$, we can define its spatial gradient (orthogonal to $u^a$) in the instantaneous rest space of the observers $\mathcal{O}_u$:

$$f_a = \nabla_a f.$$  \hfill (3.9)

In particular, to define a set of variables that characterize the fluid inhomogeneity, consider the gradients of the energy density $\mu$, pressure $p$ and expansion $\Theta$:

$$X_a = \nabla_a \mu, \quad Y_a = \nabla_a p, \quad Z_a = \nabla_a \Theta.$$  \hfill (3.10)

The following two quantities are also useful when developing a covariant theory of cosmological perturbations:

$$A = \nabla_a u^a, \quad A_a = \tilde{\nabla}_a A.$$  \hfill (3.11)

The variables defined in (3.10) have physical meaning. $X_a$ is measurable in the sense that it can be determined from virial theorem estimates, and the contribution to this quantity from luminous matter can be found by observing gradients in the number of observed sources, and then estimating the mass to light ratio [27]. A non-zero value of $X_a$ implies either a local under- or over density. However, this does not directly correspond to the quantity usually calculated. Instead, the \textit{comoving fractional density gradient} is defined as:

$$\Delta_a \equiv a \frac{X_a}{\mu}.$$  \hfill (3.12)
which represents the variation in spatial density over a fixed comoving scale. The vector can be split into a magnitude $D$ and a direction $e_a$.

\[
\Delta_a = De_a, \ e^ae_a = 1, \ e^a u_a = 0,
\]
\[
D = (\Delta_a \Delta^a)^{1/2}.
\]
(3.13)

It is the magnitude $D$ that most clearly corresponds to the usual $\delta \mu/\mu$. Note that $D$ represents a real spatial fluctuation. Similarly, an additional variable can be defined from $Z_a$:

\[
Z_a = aZ_a.
\]
(3.14)

This is the *comoving gradient of the expansion*.

### 3.4 Exact Non-linear Evolution Equations

With the basic inhomogeneity variables defined, equations are needed to describe how these quantities evolve along the fluid flow lines. Following [27], a brief sketch of the derivation of these equations for an imperfect fluid is given. The perfect fluid case was considered by Ellis and Bruni in [40] and the Newtonian equations by Ellis in [39].

#### 3.4.1 Propagation equation for the density gradient

The evolution equation for the comoving fractional density gradient is obtained by differentiating the definition of $\Delta_a$ (eq. (3.12)) with respect to proper time, using the conservation of energy and momentum equations (2.47),(2.48). The equation is then projected orthogonal to $u^a$ to give [27]:

\[
\dot{\Delta}_c = \frac{2}{3} \Theta \Delta_c - \left(1 + \frac{\rho}{\mu} \right) Z_c - (\omega^a_c + \sigma^a_c) \Delta_a - \frac{\rho}{\mu} \tilde{\nabla}_c \left( \pi_{ab} \sigma^{ab} \right)
\]
\[
+ \frac{2}{3} \Theta \left[ (\text{div } \pi)_c + \tilde{q}_c + (\omega_{cd} + \sigma_{cd} + \frac{4}{3} \Theta h_{cd}) q^d \right]
\]
\[
- \frac{\rho}{\mu} \tilde{\nabla}_c \left( (\text{div } q) + 2 \tilde{u}_c q^a \right) + \frac{1}{\mu} \left( \pi_{ab} \sigma^{ab} + (\text{div } q) + 2 \tilde{u}_a q^a \right) (\Delta_c - a \tilde{u}_c).
\]
(3.15)

This equation shows that the factors determining the time variation of $\Delta_a$ are the source term $Z_a$, non-linear terms coupling $\Delta_a$ with the shear and vorticity and imperfect fluid contributions.
3.4.2 Propagation equation for the expansion gradient

In deriving an evolution equation for the comoving gradient of the expansion, the starting point, once again, is the definition of $\mathcal{E}_a$ (3.14). This is differentiated with respect to proper time, using the Raychaudhuri (2.49) and momentum equations (2.48), and again projected orthogonal to $u^a$ [27]:

$$
\dot{\mathcal{E}}_a = a \left[ \dot{u}_c \mathcal{R} + A_c - 2 \bar{\nabla}_c (\sigma^2 - \omega^2) + \frac{3}{2} (\text{div} \pi)_c + \bar{\varrho}_c \right]
+ \frac{3}{2} \left( \omega_{cb} + \sigma_{cb} + \frac{2}{3} \Theta h_{cb} \right) q^b - \frac{1}{2} \mu \Delta_c - \left( \sigma^b_c + \omega^b_c \right) \mathcal{Z}_b - \frac{2}{3} \Theta \mathcal{E}_c,
$$

(3.16)

where

$$
\mathcal{R} = -\frac{1}{3} \Theta^2 - 2\sigma^2 + 2\omega^2 + A + \mu + \Lambda.
$$

(3.17)

The curvature scalar in the subspace $H_p$ (defined in section 2.2.1 as the instantaneous rest space of a comoving observer $\mathcal{O}_a$) of the tangent space $T_p$, orthogonal to $u^a$, is:

$$
^{(3)}R = 2 \left( -\frac{1}{3} \Theta^2 + \sigma^2 - \omega^2 + \mu + \Lambda \right).
$$

(3.18)

This allows us to write $\mathcal{R}$ as:

$$
\mathcal{R} = \frac{1}{2} (^{(3)}R - 3\sigma^2 + 3\omega^2 + A).
$$

(3.19)

If there is no vorticity ($\omega = 0$), then $^{(3)}R$ is the Ricci scalar in the hypersurfaces orthogonal to the fluid flow lines. This is discussed further in appendix B.

The equations presented in this section are exact. They are valid in absolutely any fluid flow. Although the equations presented in the previous chapter form a complete set, these propagation equations extract the details of the fluid inhomogeneity.

3.5 Almost FRW Universes

In section 3.2, the idea of a mapping between a background "smooth" spacetime, and the real "lumpy" universe was introduced. An important background model is the homogeneous and isotropic Friedmann-Robertson-Walker (FRW) universe. The fundamental requirement for any GI quantity, is that it is invariant under a choice of mapping from the homogeneous,
isotropic FRW background into the real universe. The simplest cases are scalars which are constant in the background model, or tensors which vanish everywhere in the background model. This allows the mapped quantities to be either constant or zero, respectively, and thus any chosen correspondence between the spacetimes defines the same perturbation. (The other possibilities, discussed in section 3.2.3, are not considered here). In order to know what the simplest covariantly defined quantities in a homogeneous, isotropic FRW spacetime are, we will first examine the properties of the background model.

3.5.1 Properties of FRW models

This model is characterised by a perfect fluid matter tensor, and the condition that local isotropy holds everywhere [44]. More explicitly, let \( u^a = u^a_r \) be the fluid flow vector. Local isotropy implies the vanishing of the shear and vorticity of \( u^a \), and the spatial homogeneity ensures that any spatial gradients (i.e. orthogonal to \( u^a \)) of any scalar \( f \) also vanish [27]:

\[
\sigma = \omega = 0, \quad \nabla_a f = 0, \tag{3.20}
\]

in particular, the gradients of energy density, pressure and expansion

\[
X_a \equiv \nabla_\mu, \quad Y_a \equiv \nabla_a p, \quad Z_a \equiv \nabla_a \Theta, \tag{3.21}
\]

vanish, where \( Y_a = 0 \rightarrow \dot{u}_a = 0 \). Then

\[
\mu = \mu(t), \quad p = p(t) \quad \text{and} \quad \Theta = \Theta(t) = 3H(t) \tag{3.22}
\]

depend only on the cosmic time \( t \), defined \(^2\) by the FRW fluid flow vector through \( u_a = -t_a \). FRW models are conformally flat, meaning that the Weyl tensor vanishes:

\[
C_{abcd} = 0 \Leftrightarrow E_{ab} = 0, \quad H_{ab} = 0. \tag{3.23}
\]

As mentioned above, one of the characteristics of the FRW universe is the perfect fluid form (2.31) of the energy momentum tensor. This is guaranteed by the choice of \( u^a = u^a_r \), and means that the anisotropic pressure \( \pi_{ab} \) and the energy flux \( q_a \) vanish identically. It follows that these models are completely determined by an equation of state \( p = p(\mu) \), the energy conservation equation

\[
\dot{\mu} + \Theta(\mu + p) = 0, \tag{3.24}
\]

\(^2\)up to a constant
the Raychaudhuri formula

$$\dot{\Theta} + \frac{1}{3} \Theta^2 + \frac{1}{2} (\mu + 3p) - \Lambda = 0,$$

(3.25)

and the Friedmann equation

$$\frac{1}{a^2} \dot{a}^2 + \frac{K}{a^2} = \frac{1}{3} \mu + \frac{1}{3} \Lambda,$$

(3.26)

where (3.26) is a first integral of (3.25), if \( H \neq 0 \).

Following the standard notation (e.g. [27, 6]), we define the following parameters in the background,

$$w \equiv \frac{p}{\mu}, \quad c_s^2 \equiv \frac{dp}{d\mu} = \frac{\dot{\mu}}{\dot{\mu}},$$

(3.27)

$$\Rightarrow \dot{w} = -(1 + w) \left( c_s^2 - w \right) \Theta,$$

(3.28)

where the evolution equation for \( w \) follows from the energy conservation equation, and \( c_s^2 \) is the speed of sound in the fluid.

If standard comoving coordinates are assumed, the metric of FRW is of the form [27]:

$$ds^2 = -(u_0)^2 (dx^0)^2 + h_{\alpha \beta} dx^\alpha dx^\beta$$

(3.29)

$$(u_0)^2 (dx^0)^2 = dt^2 = \ell^2 d\eta^2, \quad h_{\alpha \beta} = \ell^2 \gamma_{\alpha \beta},$$

(3.30)

where \( x^0 \) can either be \( t \) (proper time, \( u_0 = -1 \)) or \( \eta \) (conformal time, \( u_0 = -\ell \)), and \( \gamma_{\alpha \beta} \) is the metric of the 3-surfaces of constant curvature \( K = 0, \pm 1 \).

### 3.5.2 Covariantly defined GI variables

One of current the most important understandings of the nature of the universe is that at recent times it is well-represented by the standard spatially homogeneous and isotropic FRW models [44]. This belief is based on the observed high degree of isotropy of the CBR, together with a fundamental result of Ehlers, Geren and Sachs [48], which states that if a family of freely-falling observers measure self-gravitating background radiation to be everywhere exactly isotropic in the case of non-interacting matter and radiation, then the universe is exactly FRW. However, the CBR is not exactly isotropic, and on all but the largest scales, the real universe is not well
represented by a FRW model because of the inhomogeneity observed in the matter distribution. It is necessary to obtain models which can be compared with detailed observations, hence the real universe must be approximated as well as is possible. It can be shown that the EGS result remains nearly true, if the radiation is nearly isotropic. This is expressed in the almost-EGS-theorem [94], taken nowadays to establish that the universe is almost-FRW at recent times [44]. The almost-FRW model represents a universe that is FRW-like on a large scale, but allows for generic inhomogeneities on a small scale.

The two fundamental inhomogeneity variables, (see section 3.3) are the comoving fractional density gradient \( \Delta_a \), and the comoving expansion gradient \( Z_a \). The exact non-linear equations governing the evolution of these variables were shown in sections 3.4.1, and 3.4.2. We will now linearize these equations about a FRW background in order to obtain the evolution equations which are valid in an almost FRW universe. The solutions of the equations which describe this universe are a subset of the whole space of solutions of the Einstein equations, which itself contains the smaller set of exact FRW models.

3.5.3 Linearization

In order to linearize the equations, we use an order-of-magnitude notation as in [44]. Given a smallness parameter \( \epsilon \), \( \mathcal{O}[\epsilon^n] \) denotes \( \mathcal{O}(\epsilon^n) \), and \( A \approx B \) means \( A - B = \mathcal{O}[2] \). When \( A \approx 0 \) we shall regard \( A \) as vanishing when we linearize, since it is zero to the accuracy of relevant first-order calculations. We apply this scheme to almost FRW models as follows [40, 44]:

- Zero order variables \( \mu, p, \Theta \) and their time derivatives \( \dot{\mu}, \dot{p}, \dot{\Theta} \) do not vanish in the FRW background,
- First order variables are \( \dot{u}^a, \sigma_{ab}, \omega^a, q^a, \pi_{ab}, E_{ab}, H_{ab}, \Delta_a, Z_a \) and their time and space derivatives. These quantities all vanish in the FRW background.

The linearization procedure is therefore trivial. Terms of order \( \mathcal{O}[2] \), e.g. \( \sigma^b_c \Delta_b \) occurring in the exact equations are dropped. Before applying this procedure to the exact covariant equations, we will examine the various types of GI variables.
Spatial Gradients

The Stewart and Walker Lemma, together with (3.20), show that the shear and vorticity of $u^a = u^a_E$, and the spatial gradients orthogonal to it, are all GI variables. This is true in general; shear vorticity and spatial gradients which are defined with respect to any covariantly defined timelike unit vector $u^a$, coinciding with $u^a_E$ in the FRW spacetime, vanish in this background, and hence are also GI. This explains the importance of the choice of fluid 4-velocity. In particular, since the gradients (3.21) are GI, the fundamental inhomogeneity variables

$$\Delta_a = \delta_{\mu} \tilde{\nabla}_a \mu, \quad \text{and} \quad \mathcal{Z}_a = a \tilde{\nabla}_a \Theta$$

(3.31)

are also GI. The analysis of density perturbations within the covariant approach is based on these two variables and their evolution equations.

The linear equations for $\Delta_a$ (3.59) and $\mathcal{Z}_a$ (3.60) can be derived by taking spatial gradients of the linearized energy conservation (3.43) and Raychaudhuri (3.45) equations keeping only the linear terms that arise. Alternatively, the evolution equations can be obtained by directly linearizing the exact equations (3.15, 3.16). This involves the momentum equation (3.44), implying the appearance of pressure gradient terms in the equations. However, given an equation of state $p = p(\mu, s)$, where $s$ is the entropy density, the pressure gradient $Y_a$ can be substituted for in terms of the energy density and entropy gradients. The definition

$$\mathcal{E}_a = \frac{a}{p} \left( \frac{\partial p}{\partial s} \right)_{(\mu)} \tilde{\nabla}_a s,$$

(3.32)

implies

$$p \mathcal{E}_a = a Y_a - c_s^2 \dot{\mu} \Delta_a,$$

(3.33)

where $\mathcal{E}_a$ is a measure of the entropy perturbation. If the background is a standard FRW universe with vanishing bulk viscosity $\mathcal{B}$ then

$$c_s^2 = \frac{\dot{p}}{\dot{\mu}} = \left( \frac{\partial p}{\partial \mu} \right)_{(s)},$$

(3.34)

and $\dot{s} = 0$ along the flow lines.

Curvature Variables

Following Hawking [56], (also Dunsby [27]), we focus on the curvature, i.e. the Riemann tensor $R^a_{b c d}$. This consists of its trace $R_{ab} = R^c_{acb}$, the Ricci
tensor, and its trace-free part, the Weyl tensor \( C_{abcd} \), defined in (2.39), which represents the free gravitational field non-locally determined by matter. Since FRW spacetimes are conformally flat, \( C_{abcd} = 0 \), and the Weyl tensor is thus GI in any frame. This differs from the spatial gradients, shear and vorticity, since these are only GI when orthogonal to a unit timelike vector coinciding with \( u^b \), in the background. Any possible decomposition of \( C_{abcd} \) thus gives GI variables, in particular, the electric and magnetic parts of the Weyl tensor are covariant GI variables:

\[
E_{ab} = C_{abcd} u^c u^d, \quad H_{ab} = \frac{1}{2} \epsilon_{ade} C^{def} bc u^c, \tag{3.35}
\]

where \( u^c, u^d \) are completely arbitrary timelike unit vectors. \( E_{ab} \) represents a purely tidal force, but \( H_{ab} \) has no Newtonian analogue [38]. Together, these represent the free gravitational field, representing gravitational waves as well as tidal forces.

Given a unit timelike vector \( u^a \), the 3-curvature tensor \( (3) R^a_{\ bcd} \) and its trace \( (3) R_{ac} \equiv (3) R^b_{abc} = (3) R_{a bc} \) can be defined at each point. When \( u^a \) is hypersurface orthogonal, these are the Riemann and Ricci tensors of the 3-surfaces perpendicular to \( u^a \), but in general \( \omega_{ab} \neq 0 \Rightarrow (3) R_{ab} \neq 0 \), and these tensors do not then have all the usual symmetries of the Riemann and Ricci tensors. With \( u^a \) the fluid four velocity, \( (3) R_{ab} \) can be split into its trace (see also section 3.4.2)

\[
(3) R = R + 2 R_{dab} u^d u^b - 2(3H^2 + \omega^2 - \sigma^2)
= 2(\frac{1}{3} \Theta^2 + \sigma^2 - \omega^2 + \mu + \Lambda), \tag{3.36}
\]

and its GI trace-free part

\[
(3) \mathcal{R}_{ab} = -\frac{1}{3} \Theta (\sigma_{ab} + \omega_{ab}) + E_{ab} + \frac{1}{2} \pi_{ab} - \frac{2}{3} h_{ab} (\sigma^2 - \omega^2) + Q_{ab}, \tag{3.37}
\]

where

\[
Q_{ab} = \sigma_{ac} \sigma^c_{\ b} + \omega_{ac} \omega^c_{\ b} + \omega_{ac} \sigma^c_{\ b} - \omega_{bc} \sigma^c_{\ a}. \tag{3.38}
\]

These last three relations are exact, i.e. the definitions are valid in any spacetime. In an almost FRW universe, terms quadratic in the GI variables, i.e. in the shear and vorticity, will be neglected. Terms such as \( H \) will be retained to zero order (the background value.) This is an explicit example of the linearization procedure outlined above.

36
(3) $R$ is a GI variable only if the background FRW model is flat. However, a GI variable can be defined from it on taking its gradient [27]:

$$C_a = a^3 \tilde{\nabla}_a \left( (3) R \right) = -\frac{4}{3} \Theta a^2 Z_a + 2 \mu a^2 \Delta_a + 4 \sigma a^2 \Sigma_a - 4 \omega a^2 W_a.$$  \hspace{1cm} (3.39)

where

$$\Sigma_a \equiv a \tilde{\nabla}_a \sigma, \quad W_a \equiv a \tilde{\nabla}_a \omega$$ \hspace{1cm} (3.40)

are the comoving spatial gradients of the shear and vorticity respectively. $\Sigma_a$ is important in a perturbed Bianchi I model. Linearizing about a FRW model $C_a$ reduces to

$$C_a = a^3 \tilde{\nabla}_a (3) R = -4 a^2 H Z_a + 2 \mu a^2 \Delta_a.$$ \hspace{1cm} (3.41)

Although this variable is a geometrically natural quantity useful in discussing large scale density perturbations, for a barotropic perfect fluid, it is not conserved when $K \neq 0$. The following closely related quantity $\tilde{C}$ is conserved in a more general set of circumstances (i.e. $K \neq 0$), and as such is physically more significant\(^3\) [27]:

$$\tilde{C}_a = C_a - \frac{4K}{1+w} \Delta_a.$$ \hspace{1cm} (3.42)

It reduces to $C_a$ when $K = 0$.

### 3.6 Linear Evolution Equations

We now linearize the equations given in sections 3.4.1 and 3.4.2 and in chapter 2 using the procedure outlined in section 3.5.3. As mentioned, $\mu$, $p$, $\Theta$ etc. are taken to be zero order quantities, and $\Delta_a$, $Z_a$, $\sigma_{ab}$, $\omega_{ab}$, $u_a$, $\pi_{ab}$, $g_a$ etc. and their derivatives to be first order.

#### 3.6.1 Energy and momentum conservation

The linear energy and momentum conservation equations follow from the exact equations (2.47) and (2.48), where only terms up to first order have been retained:

$$\dot{\mu} + (\mu + p) \Theta + (\text{div } g) = 0,$$ \hspace{1cm} (3.43)

$$(\mu + p) \dot{u}_a + Y_a + q(a) + \frac{4}{3} \Theta q_a + (\text{div } \pi)_a = 0,$$ \hspace{1cm} (3.44)

\(^3\)The conservation of $C$ or $\tilde{C}$ implies that the evolution of the matter variable is given directly by its first order equation.
3.6.2 Linear Hydrodynamic equations

Raychaudhuri Equation

Linearizing equation (2.49) gives:

\[ \dot{\Theta} + \frac{1}{2} \Theta^2 - A + \frac{1}{2} (\mu + 3p) - \Lambda = 0. \tag{3.45} \]

This determines the evolution of the Hubble parameter H along the fluid flow lines, where \( A = \nabla_a \dot{u}^a \) is the divergence of the acceleration.

Acceleration

A propagation equation for the acceleration follows directly from the above equations of conservation of energy and momentum. This is:

\[
(\dot{u}_a) + \Theta \left( \frac{1}{3} - c_s^2 \right) u_a - \nabla_a \left( \Theta \frac{dg}{dt} \right) = \frac{c_s^2}{\mu + p} \nabla_a \nabla_b q^b - \frac{4 \Theta}{3 (\mu + p)^2} \left[ \dot{q}_a + \frac{1}{3} \Theta q_a + (\text{div } \pi)_a \right] \\
- \frac{1}{(\mu + p)^2} \left[ c_s^2 q_a + \Theta \left( \frac{7}{3} + c_s^2 \right) q_a + \frac{4}{3} \Theta^2 \left[ \frac{3}{3} + c_s^2 \right] q_a - \frac{3}{2} (\mu + 3p + 2\Lambda) q_a \right] \\
- \frac{1}{(\mu + p)^2} \left[ (1 + c_s^2) \Theta (\text{div } \pi)_a + (\text{div } \pi)_a \right]. \tag{3.46} \]

The two other hydrodynamic equations are the shear and vorticity propagation equations. These are once again derived from the corresponding exact equations (2.51) and (2.52):

Shear and vorticity propagation equations

The shear propagation equation is given by:

\[
E_{ab} + \frac{3}{2} \Theta \sigma_{ab} + \frac{c_s^2}{a(1+w)} \left[ \nabla_a \Delta_b - \frac{1}{3} h_{ab} \nabla^c \Delta_c \right] = - \frac{w}{a(1+w)} \left[ \nabla_a \mathcal{E}_b - \frac{1}{3} h_{ab} \nabla^c \mathcal{E}_c \right] \\
- \frac{1}{(\mu + p)} \left[ \nabla_a \left( \dot{q}_b + \frac{3}{3} \Theta q_b + (\text{div } \pi)_b \right) \right] \\
- \frac{1}{3} h_{ab} \nabla^c (q_c + \frac{1}{3} \Theta q_c + (\text{div } \pi)_c), \tag{3.47} \]

while the vorticity propagation equation takes the form:

\[ \omega^{(a)} - \frac{1}{3} (\text{curl } \dot{u})^a + \frac{2}{3} \Theta \omega^a = 0. \tag{3.48} \]
In terms of the vorticity tensor, this equation is:

\[
\dot{\omega}_{ab} + \frac{\dot{3}}{2} \Theta \omega_{ab} = \tilde{\nabla}_{b} \tilde{\nabla}_{a} \dot{u}_{a} \\
\quad = \frac{1}{\mu + p} \tilde{\nabla}_{b} \tilde{\nabla}_{a} p - \frac{1}{\mu + p} \tilde{\nabla}_{b} \left[ \dot{q}_{a} + \frac{\dot{3}}{2} \Theta q_{a} + (\text{div } \pi)_{a} \right] \\
\quad = \left[ c_{s}^{2} \Theta - \left( \frac{\dot{p}}{\dot{\sigma}} \right) \mu \left( \mu + p \right) \right] \omega_{ab} - \frac{1}{\mu + p} \tilde{\nabla}_{b} \left[ \dot{q}_{a} + \frac{\dot{3}}{2} \Theta q_{a} + (\text{div } \pi)_{a} \right].
\]

(3.49)

(3.50)

The \( \dot{s} \) term in the last line will often be negligible, since it is usually of first order due to bulk viscosity. The term need only be considered if the background FRW universe has bulk viscosity.

### 3.6.3 Linear constraint equations

The energy constraint, or Gauss-Codazzi relation is obtained by linearizing (3.36), as follows:

\[
\frac{1}{4} \left( \frac{\partial t}{\partial r} \right) \rho = -\frac{1}{2} \rho \rho^{2} + \frac{1}{2} \rho \left( \lambda + \frac{1}{3} \Lambda \right).
\]

(3.51)

If the perturbative terms are dropped, and \( \frac{\partial t}{\partial r} \rho \) is replaced by \( \frac{\partial t}{\partial r} \rho \), then this equation, together with the Raychaudhuri (3.45) and linearized energy conservation equations (3.43), becomes the standard set of equations for homogeneous isotropic FRW models [7].

Three other constraints follow by linearizing (2.53 - 2.55) [27]:

**Shear constraint**

\[
(\text{div } \sigma)^{a} - \frac{3}{2} \tilde{\nabla}^{a} \Theta + (\text{curl } \omega)^{a} + q^{a} = 0,
\]

(3.52)

**Vorticity constraint**

\[
\text{div } \omega = 0,
\]

(3.53)

and

**\( H \) constraint**

\[
H^{ab} + \tilde{\nabla}^{a} \omega^{b} - \text{curl } \sigma^{(ab)} = 0.
\]

(3.54)
3.6.4 Linear Maxwell-like gravitational equations

The "Maxwell-like" gravitational field equations for $E_{ab}$ and $H_{ab}$ are obtained by linearizing equations (2.57 - 2.60) [27]:

\[
\text{div } E \text{ equation }
\]
\[
(\text{div } E)^a + \frac{1}{2}(\text{div } \pi)^a - \frac{1}{3} \frac{\kappa}{a} \Delta_a + \frac{2}{3} \Theta q^a = 0,
\]  
(3.55)

\[
\text{div } H \text{ equation }
\]
\[
(\text{div } H)^a + (\mu + p)\omega^a + \frac{1}{2}(\text{curl } q)^a = 0,
\]  
(3.56)

\[
\dot{E} \text{ equation }
\]
\[
\left(\dot{E}^{(ab)} + \frac{1}{2} \pi^{(ab)}\right) - (\text{curl } H)^{ab} + \frac{1}{2} \nabla (\sigma q^b) = - \frac{1}{2} (\mu + p)\sigma^{ab} - \Theta \left(E^{ab} + \frac{1}{2} \pi^{ab}\right),
\]  
(3.57)

and

\[
\dot{H} \text{ equation }
\]
\[
\dot{H}^{(ab)} + (\text{curl } E)^{ab} - \frac{1}{2}(\text{curl } \pi)^{ab} = - \Theta H^{ab}.
\]  
(3.58)

3.6.5 Linear equations for the spatial gradients

Coupling between $\Delta_a$, $Z$

The above set of equations needs to be completed in order to investigate the evolution of density perturbations. This is obtained by using the equations for the GI inhomogeneity variables $Z_a$ and $\Delta_a$. In Ellis and Bruni [40] these equations are given for a perfect fluid. The generalized imperfect fluid equations are given in Dunsby [27] and are derived by linearizing equations (3.15) and (3.16):

\[
\dot{\Delta}_{(c)} = w\theta \Delta_c - (1 + w) Z_c + \frac{a}{\mu} \left[(\text{div } \pi)_c + \dot{q}_{(c)}\right] - \frac{a}{\mu} \nabla_c (\text{div } q),
\]  
(3.59)
\[ \tilde{Z}_{(c)} = a \left[ \hat{u}_c \mathcal{R} + A_c + \frac{3}{2} \left( (\text{div } \pi)_c + \hat{q}_{(c)} \right) \right] + 2\Theta q_{(c)} - \frac{1}{2} \Delta_c - \frac{3}{2} \Theta Z_c. \]  

(3.60)

The momentum equation (3.44) is used to find an expression for

\[ A_a = \tilde{\nabla}_a (\nabla^b \hat{u}_b) = - \frac{1}{(\mu + p)} \tilde{\nabla}^a \tilde{\nabla}_a Y^a - \frac{1}{(\mu + p)} \tilde{\nabla}^a \tilde{\nabla}_a \left[ \hat{q}^{(a)} + \frac{3}{2} \Theta q^a \right] - \frac{1}{(\mu + p)} \tilde{\nabla}^a \tilde{\nabla}_a (\text{div } \pi)^a. \]  

(3.61)

We can further manipulate this expression using the following identities (see [41], and appendix A):

\[ \tilde{\nabla}_{[a} \tilde{\nabla}_{b]} Y^c = \frac{1}{2} R_{ac} Y^c = - \frac{K}{a^2} Y_a \]  

(3.62)

and

\[ \tilde{\nabla}_{[a} \tilde{\nabla}_{b]} \hat{p} = \omega_{ab} \hat{p} = - c_s^2 (\mu + p) \Theta \omega_{ab}, \]  

(3.63)

to obtain

\[ A_a = - \frac{1}{(\mu + p)} \left( \tilde{\nabla}^2 - \frac{2K}{a^2} \right) Y_a - 2\Theta c_s^2 (\text{div } \omega)_a - \frac{1}{(\mu + p)} \tilde{\nabla}^a \tilde{\nabla}_b \left[ \hat{q}^{(b)} + \frac{3}{2} \Theta q^b \right] - \frac{1}{(\mu + p)} \tilde{\nabla}^a \tilde{\nabla}_b (\text{div } \pi)^a. \]  

(3.64)

where the notation $\tilde{\nabla}^2 \equiv \tilde{\nabla}^a \tilde{\nabla}_a$ is used.

The propagation equations (3.59), (3.60), do not contain any new dynamical information - but are implied by the equations already given. The equations extract the information about the propagation of GI variables along the fluid flow lines from the full set of equations. The relevant behaviour for structure formation can now be determined from these equations alone [27].

**Equation for the curvature gradient**

An equation for the curvature variable $\tilde{C}$ can be found by differentiating equation (3.42), and using the equations for $\Delta_a$ and $Z_a$ above:
\[
\dot{C}_a = \frac{4a^2 \Theta c_z^2}{3(1+w)} \left( \nabla^2 - \frac{2K}{a^2} \right) \Delta_a + \frac{8}{3} c_z^3 a^3 \Theta^2 \nabla^2 \omega_{ab} + \frac{4a^2 \Theta w}{3(1+w)} \left( \nabla^2 - \frac{2K}{a^2} \right) \epsilon_a \\
+ \frac{4K \Theta w}{(1+w)} \epsilon_a + \frac{4a^3}{3(\mu + p)} \nabla_a \nabla^b \left( \dot{\gamma}_a + \frac{1}{3} \Theta \gamma_a + (\text{div } \pi)_a \right) \\
+ \left[ \frac{4K a}{\mu + p} - 2a^3 \right] \nabla_a \nabla^b \gamma_b. \tag{3.65}
\]

Second order equation for \( \Delta_a \)

The dynamics of the inhomogeneity variable \( \Delta_a \) can be determined either by using the two sets of first order equations given above (coupling \( \Delta_a \) and \( Z_a \) or \( \Delta_a \) and \( \dot{C}_a \)), or by using a second order equation derived from the sets of first order equations [27]:

\[
\begin{align*}
\ddot{\Delta}_a &= \dot{A}(t) \dot{\Delta}_a - B(t) \Delta_a - c_z^2 \left( \frac{2K}{a^2} + \nabla^2 \right) \Delta_a \\
&\quad - w \left( \frac{K}{a^2} + \nabla^2 \right) \epsilon_a - 2a \Theta (1 + w) c_z^2 \nabla^b \omega_{ab} = S_a, \tag{3.66}
\end{align*}
\]

where

\[
\begin{align*}
A(t) &= \frac{1}{3} \left( 2 + 3c_z^2 - 6w \right) \Theta, \\
B(t) &= \left[ \left( \frac{1}{2} + 4w - \frac{3}{2} w - 3c_z^2 \right) \mu + (5w - 3c_z^2) \Lambda + (c_z^2 - w) \frac{12K}{a^2} \right], \\
S_a &= \frac{a}{\mu} \left\{ -3w \mu + 3\Lambda - \frac{3K}{a^2} \right\} \left[ \dot{\gamma}_a + \frac{1}{3} \Theta \gamma_a + (\text{div } \pi)_a \right] \\
&\quad + \frac{\Theta}{\mu + p} \left[ \dot{\gamma}_a + \Theta \left( \frac{\gamma_a}{3} + c_z^2 \right) q_a + \frac{4}{3} \Theta^2 \left( 3 + c_z^2 \right) q_a - \frac{2}{3} \left( \mu + 3p - 2\Lambda \right) q_a \right] \\
&\quad + \frac{\Theta (1 + c_z^2)}{\mu + p} (\text{div } \pi)_a + \frac{1}{\mu + p} (\text{div } \pi)_a \\
&\quad + \frac{1}{\mu + p} \nabla_a \left[ (\text{div } \dot{q}) + \frac{1}{3} \Theta (\text{div } q) + \nabla^b (\text{div } \pi)_b \right] \\
&\quad - \frac{2\Theta}{3(\mu + p)} \nabla_a (\text{div } q) - \frac{1}{\mu + p} \nabla_a \left[ (1 + c_z^2) \Theta (\text{div } q) + (\text{div } \dot{q}) \right]. \tag{3.69}
\end{align*}
\]
Equation (3.66) is useful in the cases where a conserved quantity does not exist. It has the form of a wave equation with extra terms due to the expansion of the universe, gravity, the spatial curvature, the cosmological constant, the divergence of the vorticity and the imperfect fluid source terms. This equation reduces to the simpler perfect fluid equation given as equation (26) in [41] when \( q_a = \pi_a = 0 \) and the equation of state is barotropic \( p = p(\mu) \Rightarrow \mathcal{E}_a = 0 \).

### 3.6.6 Consistency of the linear Equations

The full set of equations so far consists of both propagation and constraint equations. Propagation equations involve time-derivatives of the kinematic or dynamic quantities, while constraint equations involve only spatial derivatives. The consistency of these equations is an important issue. In order for the linear equations to be consistent, the constraint equations must be preserved in time along the fluid flow. This can be done by taking the time derivatives of a constraint equation, and then using the commutation relations for time and space derivatives (see appendix A) to substitute for all time derivatives occurring in the resulting equation. The result will be a new constraint equation. If this is identically satisfied, then the original constraint is preserved in time. If the new constraint equation is a completely new equation, the entire procedure is repeated until an identically satisfied equation is found. If too many non-trivial constraints arise during this process, the set of equations is found to be inconsistent. The investigation the linear constraint equations presented in this chapter may be found in [27], where the set of linear equations was shown to be consistent.

Consistency is a known property of the full Einstein equations. The linearization introduced here is consistent in that it preserves this property. This needs to be checked when considering additional physical effects such as magnetism.

### 3.7 Scalar Variables

#### 3.7.1 Local decomposition

The variable \( \Delta_a \) characterizes the spatial variation of the density orthogonal to the fluid flow. The variable contains more information than is necessary for the investigation of the growth or decay of density inhomogeneities. It is convenient to introduce a local decomposition by considering the spatial
derivative of this vector (multiplied by the scale factor $a$ for convenience). This derivative is then split into parts, analogous to the splitting of the first covariant derivative of the 4-velocity vector (2.18) \cite{44, 41}, as follows;

$$a \tilde{\nabla}_a \Delta_b \equiv \Delta_{ab} = W_{ab} + \Sigma_{(ab)} + \frac{1}{3} \Delta h_{ab}. \quad (3.70)$$

Here

$$W_{ab} \equiv \Delta_{[ab]}$$
$$= \frac{a^2}{\mu} \tilde{\nabla}_{[a} \tilde{\nabla}_{b]} \mu$$
$$= \frac{a^2}{\mu} w_{ab} \mu$$
$$= -a^2 (1 + \omega) \Theta w_{ab}, \quad (3.71)$$

using the identity (3.63) and neglecting a second order term from $\tilde{\nabla}_a \alpha$. This antisymmetric part of the spatial gradient represents the spatial variation in $\Delta_a$, in which its magnitude is preserved (i.e. rotations of this vector). $W_{ab}$ is a dipole-like density gradient and is not directly associated with matter clumping. The trace-free symmetric part

$$\Sigma_{ab} = a \tilde{\nabla}_{(b} \Delta_{a)} - \frac{1}{3} \Delta h_{ab} \quad (3.72)$$

describes the evolution of anisotropies in the universe which are pancake of cigar like structures.

Finally, the part of the density evolution that most clearly reflects the local aggregation of matter is the spatial divergence:

$$\Delta = \Delta^a_a = a \tilde{\nabla}^a \Delta_a. \quad (3.73)$$

### 3.7.2 Equations for scalar variables

We will consider equations for scalar variables that relate directly to the growth of structure in the universe. Specifically, we will concentrate on the equations for $\Delta$, since these represent the growth of density perturbations associated with the growth of matter clumping. Also, it can be shown that all the scalar variables, which can be obtained on taking divergences of the vector and tensor variables previously used, are determined to first order by $\Delta$ through the given equations see section 3.8).

We will present equations for the following GI scalar variables:
\[ \Delta = a\tilde{\nabla}_a \Delta^a, \quad Z = a\tilde{\nabla}_a Z^a, \]
\[ C = a\tilde{\nabla}_a C^a, \quad \tilde{C} = a\tilde{\nabla}_a \tilde{C}^a. \]  
(3.74)

These are derived from the analogous gradients defined previously with [7]
\[ C = -\frac{4}{3}a^2\Theta Z + 2\mu a^2 \Delta, \quad \tilde{C} = C - \frac{4\mu K}{1+w} \Delta. \]  
(3.75)

The evolution equation for \( \Delta \) is obtained by taking the divergence of (3.59):
\[ \dot{\Delta} = w\Theta \Delta - (1 + w)Z - \frac{a^2}{\mu} \tilde{\nabla}^2 (\text{div } q) + \frac{a^2}{\mu} \Theta \tilde{\nabla}^c [ (\text{div } \pi)_c + \dot{q}_c]. \]  
(3.76)

Similarly, taking the divergence of (3.60) yields the equation for \( Z \):
\[ \dot{z} = - \frac{4}{3} \Delta - \frac{2}{3} \Theta Z + a \left[ \frac{3K}{a^2} + \tilde{\nabla}^2 (\text{div } \pi)_c + (\text{div } q) + \frac{4}{3} \Theta (\text{div } q) \right] \]
\[ - \frac{a^2}{(\mu + p)} \left[ \tilde{\nabla}^2 + \frac{3K}{a^2} \right] \left[ \tilde{\nabla}^c (\text{div } \pi)_c + (\text{div } q) + \frac{4}{3} \Theta (\text{div } q) \right] \]
\[ - \frac{w}{1+w} \left[ \frac{3K}{a^2} + \tilde{\nabla}^2 \right] \mathcal{E} - \frac{1}{1+w} c^2 \left[ \frac{3K}{a^2} + \tilde{\nabla}^2 \right] \Delta. \]  
(3.77)

where \( \mathcal{E} = a\tilde{\nabla}_a C^a \).

It is sometimes practical to couple the evolution of \( \Delta \) with that of the variable \( \tilde{C} \), which is conserved in some cases. \( \tilde{C} \) satisfies:
\[ \dot{\tilde{C}} = \frac{4a^2\Theta c^2}{3(1+w)} \tilde{\nabla}^2 \Delta + \frac{4a^2\Theta w}{3(1+w)} \left[ \tilde{\nabla}^2 + \frac{3K}{a^2} \right] \mathcal{E} \]
\[ + \frac{4a^4\Theta}{3(\mu + p)} \tilde{\nabla}^2 \left[ \tilde{\nabla}^a q_a + \frac{4}{3} \Theta (\text{div } q) + \tilde{\nabla}^a (\text{div } \pi)_a \right] \]
\[ + \frac{4K a^2}{(\mu + p) - 2a^2} \tilde{\nabla}^2 (\text{div } q), \]  
(3.78)

and the corresponding coupled equation for \( \Delta \) is [27]:
\[ \dot{\Delta} = \left\{ \Theta w - \left[ \frac{(\mu + p)}{a^2} - \frac{3K}{a^2} \right] \right\} \Delta - \frac{3(1+w)}{4a^2 \Theta} \tilde{C} \]
\[ = \frac{a^2 \Theta}{\mu} \left[ \tilde{\nabla}^a q_a + \frac{4}{3} \Theta (\text{div } q) + \tilde{\nabla}^a (\text{div } \pi)_a \right] - \frac{a^2}{\mu} (\text{div } q). \]  
(3.79)

on substituting for \( Z \) using (3.75).
Second order equation

When a conserved quantity does not exist, the evolution of $\Delta$ can still be determined by a second order equation:

$$\ddot{\Delta} + \mathcal{A}(t)\dot{\Delta} + \mathcal{B}(t)\Delta - \sigma_{\phi}^2 \vec{\nabla}^2 \Delta - w \left( \vec{\nabla}^2 + \frac{2K}{\alpha^2} \right) \mathcal{E} = S \quad (3.80)$$

where $S = a\vec{\nabla}^{a}S_{a}$, and $\mathcal{A}(t)$, $\mathcal{B}(t)$ are given by (3.67, 3.68). This equation follows from either of the above pairs of equations coupling $\Delta$ with either $\mathcal{Z}$ or $\mathcal{C}$.

3.8 Local Classification of Perturbations

Because the equations have been obtained by linearization, their solution can be regarded as consisting additively of different parts, each of which is separately a solution of these linear equations. It is a well known fact that equation (3.50) shows that the rotational perturbations evolve independently of any variables other than $\omega_{ab}$ [7]. In general, these variables affect the evolution of other quantities. Since the equations are linear, this means that in the general case all the vectorial and tensorial variables have a rotational "vector" contribution, which can be regarded as a known source term.

As is shown in [27], there will be a non-trivial scalar contribution to the perturbed model if and only if $\Delta \neq 0$. The vector contribution to the perturbed model is non-trivial if and only if $\omega_{ab} \neq 0$. In the third case of tensor (gravitational wave) perturbations, there are no non-trivial perturbations if $\omega_{ab} = 0 \Rightarrow E_{ab} = H_{ab} = 0$. Gravitational waves are represented by those parts of the Weyl tensor components $E_{ab}$ and $H_{ab}$ which do not arise from rotational ("vector") and density ("scalar") perturbations, i.e. by their transverse traceless (TT) parts satisfying $\vec{\nabla}^{a}E_{ab} = 0$, and $\vec{\nabla}^{a}H_{ab} = 0$ [27].

In deriving these results, the matter is assumed to be described by a barotropic perfect fluid with $p = p(\mu)$.

3.9 Solutions and Harmonic Components

Since rotational perturbations evolve independently (see equation (3.50)), vorticity terms can be considered as known source terms in the vector and tensor equations, producing a rotational mode in the corresponding variable. In considering scalar equations, when the full divergence of the equations is
taken, the gravitational (TT tensor) and rotational modes disappear, since
\[ \nabla^a \nabla^b \omega_{ab} = 0. \]
This means that even if the vorticity does not vanish, for the purposes of the initial value problem the scalar equations can be solved as if \( \omega_{ab} = 0 \) [27].

To actually solve the equation, it is standard to assume that the time and spatial dependence in each variable are separable, decomposing each quantity in spatial harmonics. This separates the differential equations for time variation of the perturbations as a whole into separate time variation equation for each component of the spatial variation, characterized by comoving wave number. It is convenient to use a set of harmonics \( Q \) which are covariantly constants, i.e. \( Q = 0 \), defined as eigenfunctions of the covariant Laplace-Beltrami operator
\[
\tilde{\nabla}^2 Q = -\frac{k^2}{a^2} Q, \tag{3.81}
\]
where \( Q \) represents any of the scalar, vector or tensor harmonics \( Q^{(0)}, Q_a^{(1)} \) and \( Q_{ab}^{(2)}. \) Using the \( Q \)'s, the equations provide standard results, discussed in [40, 41, 42]. The use of a comoving wave number \( k \) to characterize each spatial variation conveniently encompasses the idea of a comoving wavenumber for the matter inhomogeneities. In particular, the use of harmonics allow for the definition of a "long wavelength limit", which enables one to consider solutions for those wavelengths \( \lambda = (2\pi a)/k \) larger than the Hubble scale, so that \( H^{-1}/\lambda << 1 \). In this case, in a \( K = 0 \) universe, \( \bar{C} = C \) and all the Laplacian terms on the right hand side of (3.78) can be neglected. The curvature scalar is preserved in this limit, and equation (3.76) can be integrated directly. For \( K \neq 0 \), the variable \( \bar{C} \) is conserved under appropriate circumstances (see [41]).

The general solution of (3.80) for a perfect fluid in a flat \( K = 0 \) background was given in [6]. Here we look at two particular perfect fluid solutions in a flat \( K = 0 \) background [44, 27].

### 3.9.1 Solutions

**Dust**

Here \( p = 0 \Rightarrow w = c_s^2 = 0 \), and the equation (3.80) takes the form:
\[
\ddot{\Delta} + \frac{2}{3} \Theta \dot{\Delta} - \frac{4}{9} \Delta = 0. \tag{3.82}
\]
This has two solutions, proportional to
\[ \Delta_+ \propto t^{\frac{2}{3}}, \quad \Delta_- \propto t^{-1}; \] (3.33)
representing modes of increasing and decreasing inhomogeneity respectively. As for radiation, these are known as growing or decaying modes.

3.9.2 Radiation
Here \( p = \frac{1}{3} \mu \Rightarrow w = c_s^2 = \frac{1}{3} \). The equation to solve becomes:
\[ \ddot{\Delta} + \frac{1}{3} \Theta \dot{\Delta} - \left( \frac{2}{3} \mu - \frac{1}{3} \frac{\dot{c_s}^2}{\dot{\Omega}} \right) \Delta = 0. \] (3.84)
In the long wavelength limit, the solutions to this equation are [44, 27]:
\[ \Delta_+ \propto t, \quad \Delta_- \propto t^{-\frac{1}{2}}; \] (3.85)
In the above solutions, \( \Delta_+ \) and \( \Delta_- \) represent growing and decaying modes respectively.
Chapter 4

Dynamical Systems Theory

4.1 General Introduction

Dynamical systems theory is used to study physical systems that evolve in time, but can be applied to systems which evolve with respect to any independent variable. Instead of looking for a particular solution to a set of differential equations, dynamical systems theory looks at the qualitative properties of the set of all solutions to a system. It provides a way to understand the space of solutions without necessarily being able to solve the equations themselves. In particular, dynamical systems analyses provide a relatively easy way to study the asymptotic stability behaviour of solutions to a system of equations.

A dynamical system usually takes the form of a set of differential equations describing the evolution of the physical system. The physical variables describing the state of the system can be represented as a vector \( \vec{x} \) in the state space \( X \) of the system [111] (the space describing all the possible states of the system). The state space may be finite dimensional (e.g. \( \mathbb{R}^n \)) or infinite dimensional (e.g. a function space). We only consider finite dimensional state spaces in this thesis. The evolution of the system must be described by some law which determines the state of the system at any "time", given the state at some initial time. \(^1\) The general situation can be represented as follows:

\[
\vec{x}' = \vec{f}(\vec{x}, t),
\]

where \( ' \) denotes differentiation with respect to time, or the independent

\(^1\)The evolution may also be described with respect to a different independent variable i.e. it may not explicitly involve time evolution.
variable. If the evolution of the system does not depend on the independent variable, the system is said to be autonomous, and can be represented as:

\[ \dot{x} = \bar{f}(x), \quad (4.2) \]

where \( \bar{f} : X \to X \). In the case where \( X = \mathbb{R}^n \), (4.2) represents a system of \( n \) ordinary differential equations. A solution of this system is a curve in \( \mathbb{R}^n \) to which the vector field \( \bar{f} \) is always tangent i.e. a function \( \psi : \mathbb{R} \to \mathbb{R}^n \) which satisfies

\[ \psi'(t) = \bar{f}(\psi(t)), \quad (4.3) \]

for all \( t \in \mathbb{R} \), or possibly in a finite interval.

If solutions of a set of differential equations (4.2) can be extended for all times, the flow of the differential equations can be defined as follows [111]: Let \( \psi_a(t) \) be the unique maximal solution which satisfies \( \psi_a(0) = a \). The flow is defined to be the one-parameter family of maps \( \{\phi_t\}_{t \in \mathbb{R}} \) of \( \mathbb{R}^n \) into itself such that

\[ \phi_t(a) = \psi_a(t), \quad \text{for all } a \in \mathbb{R}^n. \quad (4.4) \]

### 4.2 Analysis

#### 4.2.1 Phase spaces and trajectories

A very useful aspect of analysing a set of differential equations as a dynamical system, is that the solution can easily be presented graphically. While this is simply a useful tool for equations which admit exact solutions, it becomes a necessity for equations which can only be treated numerically. For a one-dimensional dynamical system, the solution can be plotted as a function of time, but this becomes difficult for higher dimensional systems. In general, for an \( n \)-dimensional system, the solution must be plotted in an \( (n + 1) \)-dimensional space. This difficulty can be partially overcome by the use of phase spaces.

A phase space is the space obtained by projecting the time axis out of the solutions. Although information about how points move along the projected solution curve as functions of time is lost, the phase space still provides a useful analytical tool. For example, consider a two-dimensional system for \( x \) and \( y \) as functions of time [75]. By projecting the solutions of the system onto the \( xy \)-plane, the solutions can be visualized as curves in a two-dimensional space, thereby reducing the number of dimensions needed to plot the solutions by one. The projected curve, known as an "orbit", or
Figure 4.1: Some orbits in the $x - y$ plane for the non-autonomous system described by $\dot{x} = y$, $\dot{y} = x^2 - t$.

Figure 4.2: Some orbits in the $x - y$ plane for the autonomous system described by $\dot{x} = y$, $\dot{y} = x^2$.

"trajectory" in the $xy$-plane, does not correspond to the actual motion of a solution to the system. In this case, the $xy$-plane is an example of a "phase space" of the equations. This technique is not useful for non-autonomous systems as the orbits tend to cross one another and become indecipherable as in figure 4.1. For autonomous systems, the solutions at different values of the initial time are time translations of each other, and their projections 'pile up' on the same orbits, not crossing each other in the $x - y$ plane [75], see figure 4.2.

4.2.2 Equilibrium Points

Systems which are linear with respect to $\vec{x}$ can be written as

$$\vec{x}' = \vec{A}\vec{x}.$$  

(4.5)

Linear systems include those where $\vec{A}$ is a function of the independent variable as well as those with

$$\vec{x}' = \vec{A}(t)\vec{x} + \vec{g}(t),$$  

(4.6)

where $\vec{g}$ is any vector. We will only consider the simpler systems, which can be written as (4.5), above. A solution to these systems can be written as:

$$\vec{x}(t) = \sum_{i=1}^{n} a_i e^{\lambda_i (t - t_0)} \vec{v}_i,$$  

(4.7)

where $a_i$ are constants, $\lambda_i$ the eigenvalues of the matrix $\vec{A}$ and $\vec{v}_i$ the corresponding eigenvectors. The eigenvalues $\lambda_i$ may be imaginary.
In order to analyse the dynamical system, we look for stationary, or equilibrium points. These may also be referred to as critical, or fixed points. These are points \( \bar{x}_0 \) where the system is not evolving - i.e.

\[
\bar{x}_0' = f(\bar{x}_0) = 0,
\]  

(4.8)

and are exact solutions of the equations represented by (4.5).

The behaviour of the solutions at the stationary points depends on the eigenvalues \( \lambda_i \) of the matrix \( \bar{A} \) at each point. Furthermore, the stationary points are also classified based on the signs of the real parts of the eigenvalues.

Classification in two dimensions

In linear, two dimensional autonomous systems, the only equilibrium point is at the origin. Three cases can be identified. All the following examples, except case 3b, have the general solution (4.7).

1. Real and distinct eigenvalues (Figures 4.3 - 4.5) \(^2\)

   The behaviour of the solutions depends on the sign of the eigenvalues.

   (a) **Node Source**: \( 0 < \lambda_1 < \lambda_2 \)

   With both eigenvalues positive, all orbits tend to infinity as \( t \to \infty \) and to the origin as \( t \to -\infty \).

   (b) **Saddle**: \( \lambda_2 < 0 < \lambda_1 \)

   Orbits in the direction of positive and negative multiples of the eigenvector \( \bar{v}_2 \), corresponding to the eigenvalue \( \lambda_2 \), will tend towards the origin as \( t \to \infty \) and orbits in the direction of the other eigenvector \( \bar{v}_1 \), will tend towards infinity as \( t \to \infty \). Other orbits are superpositions of these orbits. This type of equilibrium point is sometimes referred to as hyperbolic.

   (c) **Node Sink**: \( \lambda_1 < \lambda_2 < 0 \)

   All orbits tend towards the origin as \( t \to \infty \), and towards infinity as \( t \to -\infty \).

2. Complex eigenvalues (Figures 4.6 - 4.8)

   These take the form \( \lambda = \alpha \pm i\beta \), where \( \alpha \) and \( \beta \neq 0 \) are real numbers.

---

\(^2\)All figures in this chapter courtesy of Ulf Nilsson [75].
Figure 4.3: The behavior of trajectories around a node source with $\lambda_1 \neq \lambda_2$ strictly positive.

Figure 4.4: The behavior of trajectories around a node saddle with $\lambda_1 < 0 < \lambda_2$.

Figure 4.5: The behavior of trajectories around a node sink with $\lambda_1 \neq \lambda_2$ strictly negative.

Figure 4.6: The behavior of trajectories around a spiral sink with $\lambda$ imaginary and $\alpha < 0$.

Figure 4.7: The behavior of trajectories around a center with $\lambda$ purely imaginary since $\alpha = 0$.

Figure 4.8: The behavior of trajectories around a spiral source with $\lambda$ imaginary and $\alpha > 0$.

The behaviour of solutions depends on the sign of the real part of the eigenvalue.

(a) **Spiral Sink (attractive focus) $\alpha < 0$**
All orbits turn around the origin at a speed of $\beta$ radians per unit time and approach the origin as $t \to \infty$.

(b) **Center $\alpha = 0$**
Orbits are ellipses around the origin.

(c) **Spiral Source (repulsive focus) $\alpha > 0$**
Orbits tend towards infinity as $t \to \infty$, spiralling outwards.
The degenerate case, with two linearly independent eigenvectors, and $\lambda_1 = \lambda_2 = \lambda$ strictly positive.

Figure 4.9:

The degenerate case, showing a node sink with $\lambda_1 = \lambda_2 = \lambda$ strictly negative.

Figure 4.10:

The behavior of trajectories around a line of sources, occurring when $\lambda_1 > 0, \lambda_2 = 0$.

Figure 4.11:

3. Repeated (double) eigenvalues (Figures 4.9 - 4.11).

This occurs when $\lambda_1 = \lambda_2 = \lambda$. There are two subcases:

(a) Infinitely many eigenvectors.

In this case, two eigenvectors can be chosen to be a basis of $\mathbb{R}^2$.

The orbits tend to the origin or to infinity depending on the sign of the eigenvalues, just like the node sink and node source, above.

(b) One eigenvector

In this case, the general solution is not of the form (4.7), but can be written as [75]:

\[
\begin{align*}
    x &= a_1 A_1 e^{\lambda t} + a_2 (A_1 + A_2 t) e^{\lambda t}, \\
    y &= a_1 B_1 e^{\lambda t} + a_2 (B_1 + B_2 t) e^{\lambda t},
\end{align*}
\]

(4.9)

where the $A's$ and $B's$ are determined by the linear differential equations (4.5), and the $a_{1,2}$ are the usual arbitrary constants.

There is also the trivial case where both eigenvalues are zero, resulting in a dynamically uninteresting solution. When just one eigenvalue is zero, the solution is a line of equilibrium points. Depending on the sign of the remaining eigenvalue, the points will be node sinks or node sources.

**Classification in three dimensions**

This follows naturally from the classification in two dimensions. The three eigenvalues can be written $\lambda_1, \lambda_2, \lambda_3$, one of which is necessarily real. The
equilibrium points can again be classified as node sinks, sources and saddles (for \( \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R} \)), or as spiral saddles, sinks and sources. A zero eigenvalue gives rise to a line of points which can be classified as in two dimensions. Examples are shown in figures 4.12 - 4.14.

### 4.2.3 Invariant sets

The concept of an invariant set is one of the most important concepts of dynamical systems. The definition of an invariant set is as follows [111]:

A set \( S \subset \mathbb{R}^n \) is an **invariant set** of the flow \( \phi_t \) on \( \mathbb{R}^n \), or of the corresponding differential equations e.g. (4.2), if, for all \( \vec{x} \in S \) and for all \( t \in \mathbb{R} \), \( \phi_t(\vec{x}) \in S \).

Lower dimensional invariant sets can be thought of as describing a restricted class of physical systems, satisfying some special property, or whose evolution is restricted in some way. Lower dimensional invariant sets often describe the asymptotic behaviour of general classes of solutions as \( t \to +\infty \).
4.2.4 Non-linear systems

In general, the non-linear (autonomous) system defined by:

\[
\dot{x} = f(x)
\]

(4.10)
can not be solved exactly. However, it is possible to show that the local behaviour of the non-linear system (4.10) near a hyperbolic equilibrium point \(x_0\) is qualitatively determined by the behaviour of the linear system (4.5) near the origin when the matrix \(A\) is equal to \(Df(x_0)\), which is the derivative of \(f\) at \(x_0\). This result is known as the Hartman-Grobman theorem [75, 82]. The linear system derived from (4.10), called the linearization of (4.10) at \(x_0\), can be written as:

\[
\dot{x} = Df(x_0) \cdot x,
\]

(4.11)
where

\[
f = \begin{pmatrix}
    f_1 \\
    f_2 \\
    \vdots \\
    f_n
\end{pmatrix}
\]

(4.12)
and \(f_1, \ldots, f_n\) are ordinary functions of \(x\) and

\[
Df(x_0) = \begin{pmatrix}
    \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\
    \vdots & \ddots & \vdots \\
    \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n}
\end{pmatrix},
\]

(4.13)
where all partial derivatives are evaluated at \(x = x_0\).

This is essentially a Taylor expansion of the functions \(f_1, \ldots, f_n\) around the origin (since this is the usual equilibrium point of a linear system). If the equilibrium point is not at the origin, new variables can be introduced so that the linear system, in the new variable, has its equilibrium point at the origin e.g. \(\xi = x - x_0\). The Hartman-Grobman theorem is only formulated for hyperbolic equilibrium points, i.e. equilibrium points where all real parts of the eigenvalues of \(Df(x_0)\) are non-zero. These cases occur quite often in general relativity. If only one eigenvalue is equal to zero, then the equilibrium set is called normally hyperbolic, and the linearization is still expected to give the correct qualitative dynamics [75].

56
Chapter 5

Dynamical Systems Analysis of Simple Perturbation Examples

5.1 Introduction

As described in chapter 3, the perturbation approach for studying cosmological density inhomogeneities starts from the exact non-linear equations for the inhomogeneity variables. These equations can then in principle be linearized about any FRW or non-tilted Bianchi model. The background model is described by $n$ dimensionless variables, denoted by $\vec{y} \in \mathbb{R}^n$ where the variables, $\vec{y}$ satisfy a differential equation of the form [112]:

$$\frac{d\vec{y}}{d\tau} = \vec{f}(\vec{y}),$$

(5.1)

$\tau$ is the dimensionless time variable defined in conjunction with the scale factor as

$$a = a_0 e^\tau,$$

(5.2)

and $a_0$ is the value of the scale factor at some arbitrary reference time. The dimensionless inhomogeneity variables are denoted by $\vec{U} \in \mathbb{R}^m$. Their evolution equations are written in the form

$$\frac{d\vec{U}}{d\tau} = g(\vec{U}, \vec{y}).$$

(5.3)
The two equations, (5.1), (5.3), form a coupled system of differential equations. The state of the system at any time can be represented by a state vector \((\mathbf{y}, \mathbf{U}) \in \mathbb{R}^{n+m}\). Dynamical systems techniques can then be used to investigate the stability of the perturbations, which these equations describe, without any need to solve explicitly for the background variables. The evolution of perturbations is then represented by trajectories in the phase space of the dynamical system, and the stability of such perturbations is linked to the presence and nature of the stationary (fixed) points of the system. This approach has been applied to the flat and open FRW models [108, 9, 8] and to locally rotaionally symmetric (LRS) Bianchi I models [28]. An introduction to this approach, covering both the FRW and the Bianchi I models is given by Dunsby in [112].

### 5.2 Peturbations of FRW Universes

The key perturbation equations were derived in chapter 3. Recall, the main inhomogeneity variables are the comoving gradients of density \(\Delta_a\) (3.12), and expansion \(\mathcal{Z}_a\) (3.14), or the corresponding scalar variables \(\Delta = a\nabla^a \Delta_a\) and \(\mathcal{Z} = a\nabla^a \mathcal{Z}_a\). (see (3.74)).

As usual, the scalar density perturbation variable satisfies a second order (in some time variable) differential equation. It was shown in section 3.9 that, after harmonic analysis, and assuming adiabatic perturbations, the evolution of the harmonic component \(\Delta_{(k)}\) of the density perturbation is given by a homogeneous ordinary differential equation:

\[
\dot{\Delta}_{(k)} + \alpha(t) \Delta_{(k)} + \beta(t) \Delta_{(k)} = 0,
\]

where \(t\) here is proper time, the dot indicates a derivative with respect to \(t\) and subscript \(k\) is ignored from now on. The coefficients \(\alpha\) and \(\beta\) are functions given by the background dynamics. In general, their time dependence cannot be explicitly determined. It is useful to think of these coefficients as known functions of the background parameters, and to include, as part of the dynamical system, evolution equations for these background parameters in order to obtain an autonomous system [8]. Woszczyna [108] used the density parameter \(\Omega\) as this parameter, where

\[
\Omega = \frac{\mu}{3H^2},
\]

where \(H = \frac{1}{3} \Theta\) is the Hubble parameter.
The FRW models, as before, are determined by an equation of state, the energy conservation equation (3.24) and the Friedmann equation (3.26), which, for an FRW model, reads [112],
\[
\mu - 3H^2 = \frac{1}{2} (3) R = \frac{3K}{a^2}. \tag{5.6}
\]
In terms of \( \Omega \), (5.6) can be written as:
\[
\Omega - 1 = \frac{K}{a^2 H^2}. \tag{5.7}
\]
The evolution of \( \Omega \) and \( w = p/\mu \) in an FRW model is given by:
\[
\dot{\Omega} = \Omega(\Omega - 1)(1 + 3w)H, \tag{5.8}
\]
\[
\dot{w} = -3(1 + w) \left( c_s^2 - w \right) H. \tag{5.9}
\]

### 5.3 FRW Perfect fluid example

To get the equations for the scalar variables, the exact equations (3.15) and (3.16) can be linearized, after dropping imperfect fluid terms, to obtain [112]:
\[
\dot{\Delta}_a = 3wH \Delta_a - (1 + w) Z_a, \tag{5.10}
\]
\[
Z_a = -\frac{2}{3} \Theta Z_a - \left[ \frac{1}{2} \mu + c_s^2 \frac{\mu - 3H^2}{1 + w} \right] \Delta_a + a A_c.
\]
Taking the divergence of these equations gives the corresponding scalar propagation equations:
\[
\dot{\Delta} = 3wH \Delta - (1 + w) Z, \tag{5.11}
\]
\[
\dot{Z} = -2H Z - \left[ \frac{1}{2} \mu + c_s^2 \frac{\mu - 3H^2}{1 + w} \right] - \frac{c_s^2}{(1 + w)} \hat{\nabla}^2 \Delta. \tag{5.12}
\]
These can also be obtained from the more general scalar propagation equations for an imperfect fluid, (3.76) and (3.77), by dropping the terms which are zero for a perfect fluid, and using the Friedmann equation (5.6) above to substitute for \( \frac{3K}{a^2} \). In the following discussion, a vanishing cosmological constant (\( \Lambda = 0 \)) is assumed.
Equations (5.11) and (5.12) can be combined to give a single, second order differential equation in $\Delta$ alone, and on harmonically analysing we obtain:

$$\ddot{\Delta} + \alpha(t)\dot{\Delta} + \beta(t)\Delta = 0,$$

(5.13)

where $\alpha(t)$ is given by

$$\alpha(t) = (2 + 3c_s^2 - 6w)H,$$

(5.14)

and

$$\beta = -3H^2 \left( \left( \frac{1}{2} + 4w - \frac{3}{2}w^2 - 3c_s^2 \right) \Omega \right) - \frac{12K}{a^4} (c_s^2 - w) + c_s^2 k^2 a^2.$$

(5.15)

### 5.3.1 Flat models

**Dust**

Recall from chapter 3 that, in the case of pressure-free matter,

$$w = 0, \quad c_s^2 = 0.$$

(5.16)

This simplifies the coefficients (5.14) and (5.15) considerably, resulting in a differential equation of the form:

$$\ddot{\Delta} + 2H\dot{\Delta} - \frac{3}{2}H^2\Omega\Delta = 0,$$

(5.17)

which is independent of $k$. For a flat FRW model,

$$\Omega = 1, \quad H = \frac{\dot{a}}{a}, \quad K = 0,$$

(5.18)

and the general solution of this differential equation is given by [112]:

$$\Delta = c_+ t^{\frac{2}{3}} + c_- t^{-1},$$

(5.19)

showing the usual growing and decaying behaviour.
General matter (radiation)

In the case of non-zero pressure, we restrict our attention to an equation of state \( p = w \rho \) where \( w \) is a constant, the second order differential equation (5.13) has the form [112]:

\[
\ddot{\Delta} + (2 - 3w)H \dot{\Delta} + H^2 \left[ -\frac{3}{2}(-3w^2 + 2w + 1) + c_s^2 \frac{k^2}{a^2 H^2} \right] \Delta = 0
\]  

(5.20)

and the dust solution given above is recovered if \( w = 0 \). For other values of \( w \), the general solution of (5.20) can be written in terms of Bessel functions (see references in [112]). However in the case where wavelengths \( \lambda = 2\pi a/k \) are much larger than the Hubble radius, characterized by

\[
\frac{k}{aH} << 1,
\]  

(5.21)

the analysis becomes very simple. In this case, equation (5.20) reduces to

\[
\ddot{\Delta} + (2 - 3w)H \dot{\Delta} - \frac{3}{2}H^2(-3w^2 + 2w + 1)\Delta = 0
\]  

(5.22)

which is independent of the wave number \( k \). The general solution, in terms of \( t \), is [112]

\[
\Delta = c_+ t^{\frac{2(3w+1)}{3(w+1)}} + c_- t^{\frac{1}{w+1}}.
\]  

(5.23)

For radiation (\( w = \frac{1}{3} \)), we obtain

\[
\Delta = c_+ t + c_- t^{-\frac{1}{2}}.
\]  

(5.24)

5.3.2 Open models

In the case of open FRW backgrounds, it is difficult to treat the coefficients of (5.20) as functions of \( t \) to obtain explicit solutions. In this case, the coefficients of (5.13) can be expressed as functions of the density parameter \( \Omega \), and the dynamical system is closed by including the evolution equation for this variable [108, 112]. For an open model \( K = -1 \). To obtain a differential equation with constant coefficients, we introduce the dimensionless time variable \( \tau \) (5.2), written as [112, 8]

\[
\tau = \ln \left( \frac{a}{a_0} \right), \quad \frac{d\tau}{dt} = H = \frac{1}{3} \Theta.
\]  

(5.25)
The basic differential equation (5.13), is given by in terms of \( \tau \) as [112]:

\[
\Delta'' + \psi(\Omega)\Delta' + \xi(\Omega)\Delta = 0, 
\]

where

\[
\psi(\Omega) = (1 - 3w) - \frac{1}{2}(3w + 1)\Omega, 
\]

\[
\xi(\Omega) = -\frac{3}{2}(1 - w)(3w - 1)\Omega + k^2w(1 - \Omega). 
\]

The coefficients are obtained using the substitution \( w = c_t^2 \) as well as the Raychaudhuri equation (3.45). Equation (5.7) is used to express these coefficients in terms of \( \Omega \).

It is possible to reduce the order of the system by writing (5.26) as a first order equation. This is accomplished by means of the substitution

\[
\mathcal{U} = \frac{\Delta'}{\Delta}, 
\]

which leads to

\[
\mathcal{U}' = -\mathcal{U}^2 - \psi(\Omega)\mathcal{U} - \xi(\Omega). 
\]

In order to close the system of equations, the evolution of \( \Omega \) (5.8) must be written in terms of the time variable \( \tau \) as follows:

\[
\Omega' = -(3w + 1)(1 - \Omega)\Omega. 
\]

The above equations (5.30) and (5.31) correspond to the general equations (5.3) and (5.1) respectively. They form an autonomous system of equations for \((\Omega, \mathcal{U})\) which describes the evolution of density perturbations in an open FRW universe.

The usual way to write a second order differential equation such as (5.26) as a first order equation is to use \((X, Y) = (\Delta, \Delta')\) as variables. Given the definition (5.29), \(\mathcal{U}\) should be regarded as an angular variable, equal to \(\tan \theta\), where \(\theta\) is the usual polar angle in the \((\Delta, \Delta')\) plane, with \(0 \leq \theta \leq 2\pi\) [112]. Two separate spaces can be considered, the subset \(D\) of \(\mathbb{R}^2\) defined by

\[
0 \leq \Omega \leq 1, \quad -\infty < \mathcal{U} < +\infty, 
\]

and the cylinder \([0, 1] \times S^1\) defined by

\[
0 \leq \Omega \leq 1, \quad -\frac{\pi}{2} \leq \theta < \frac{3\pi}{2}, 
\]
with $-\frac{\pi}{2}$ and $\frac{3\pi}{2}$ identified. The set $D$, under the 2 - 1 map

$$(\Omega, \theta) \rightarrow (\Omega, \tan \theta),$$

(5.34)

is the image of the two halves of the above cylinder that are given by $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ and $\frac{\pi}{2} < \theta < \frac{3\pi}{2}$. As a point makes one revolution on the cylinder, the corresponding image point traverses the strip $D$ twice. In principle, $\theta$ should be used as the variable, and a differential equation for $\theta'$ should be derived which corresponds to (5.30). In practice it is more convenient to use (5.30) directly, and to draw the phase portraits in the subset of $\mathbb{R}^2$ defined by (5.32). Conceptually, however, we regard the state (phase) space as the cylinder given by (5.33). This set is compact, and thus there will be a future attractor for all values of $w$ and $k$ which will determine the nature of the perturbations. The invariant sets $\Omega = 0$ and $\Omega = 1$ (see (5.31)) are circles on the cylinder [112].

Because of the above interpretation, $\mathcal{U}$ can be thought of as the phase of the perturbation. If $\mathcal{U} > 0$ during the evolution, it follows from (5.29) that the inhomogeneity measured by $\Delta$ is increasing at that time (since both $\Delta'$, $\Delta > 0$, or $\Delta'$, $\Delta < 0$) while if $\mathcal{U} < 0$, the perturbations are decaying. If an orbit of the differential equation (5.30) is asymptotic to an equilibrium point, then the perturbation approaches a stationary state, decaying to zero if $\mathcal{U} < 0$ and growing if $\mathcal{U} > 0$. If an orbit is asymptotic to a periodic orbit on the cylinder, then the perturbation oscillates indefinitely, i.e. it propagates as sound waves.

**Stability Analysis**

Since $\Omega$ is a monotonic function for $0 < \Omega < 1$, the qualitative analysis of the system (5.30) - (5.31) on the invariant set (5.32) is simple. The equilibrium points are given by [112]

$$\begin{align*}
\Omega &= 1, \quad \mathcal{U} = -\frac{3}{2}(1 - w), \quad (3w + 1), \\
\Omega &= 0, \quad \mathcal{U} = -\frac{1}{2}(1 - 3w) \left[ 1 \pm \sqrt{1 - \left( \frac{k}{k_{\text{crit}}} \right)^2} \right],
\end{align*}$$

(5.35)

where

$$k_{\text{crit}}^2 = \frac{(1 - 3w)^2}{4w}.$$  

(5.36)

The wavenumber $k$ satisfies $k^2 > 1$ [9]. If equilibrium points with $\Omega = 0$ are to exist, $k^2 < k_{\text{crit}}^2$, which requires $k_{\text{crit}} > 1$. It can be seen from (5.36) that
this is equivalent to:

\[ 0 < w < \frac{1}{\Omega} \text{ or } w > 1. \]  

(5.37)

If \( w = 0 \), the equilibrium points with \( \Omega = 0 \) are

\[ \Omega = 0, \quad \dot{\mathcal{U}} = 0 \text{ or } -1, \]  

(5.38)

independently of the value of \( k \). The phase portraits are shown in figures 5.1 and 5.2. There are two topologically inequivalent cases. Again, the state space is interpreted as the cylinder (5.33). When there are no equilibrium points with \( \Omega = 0 \) (figure 5.2), the invariant set \( \Omega = 0 \) is a periodic orbit, and the system evolves towards this, resulting in perturbations which oscillate as sound waves. In the case where there are two equilibrium points with \( \Omega = 0 \) (figure 5.1), one is a saddle and the other is a stable node (future attractor). The physical interpretation, as well as the stability behaviour of the model then depends on the value of \( \mathcal{U} \) at these points. From the definition (5.29), when \( \mathcal{U} = 0 \), the density perturbation \( \Delta \) evolves to a constant value. When \( \mathcal{U} < 0 \) the inhomogeneity is decreasing, and the density perturbation decays, as is shown in figure 5.1.

Thus, the open model is either stable or marginally stable with respect to linear inhomogeneities depending on whether \( \mathcal{U} < 0 \) or \( \mathcal{U} = 0 \).
Figure 5.2: Phase space for open models, showing the case where the future attractor is the periodic orbit $\Omega = 0$

5.4 FRW 2-fluid approach

In this section we consider relativistic perturbations of FRW models containing both dust and radiation coupled only through gravity. Following Bruni and Piotrkowska [8], and considering only the total density perturbation, we examine the stability properties of these models. We assume adiabatic perturbations. If we were to discuss the full dynamics of this situation, we would have to represent the velocities of each fluid separately. In the discussion that follows, this is not necessary because we are only considering the equations representing the total density evolution, and not that of the partial densities. As before, the system is given by the equation describing the evolution of the density perturbations (5.13),

$$\dot{\Delta} + \alpha \Delta + \beta \Delta = 0. \quad (5.39)$$

where the coefficients are defined as in (5.14) and (5.15). In terms of the time variable $\tau$, this equation becomes:

$$\Delta'' + \psi(w, \Omega)\Delta' + \xi(w, \Omega) = 0, \quad (5.40)$$
with the coefficients given by [8] \(^1\)

\[
\psi(w, \Omega) = \frac{\dot{H}}{H^2} + \frac{1}{H} \alpha(w, \Omega)
\]

\[
= -1 - \frac{1}{2} (1 + 3w)\Omega + \frac{1}{H} \alpha(w, \Omega), \tag{5.41}
\]

\[
\xi(w, \Omega) = \frac{1}{H^2} \beta(w, \Omega), \tag{5.42}
\]

where the second step in (5.41) is given by the Raychaudhuri equation (3.45) with \(\Lambda = 0\).

As before, the dynamical system also includes evolution equations for \(\Omega\) (5.31),

\[
\Omega' = (1 + 3w)(\Omega - 1)\Omega, \tag{5.43}
\]

and for \(w\) (5.9), which, in terms of \(\tau\), is written as:

\[
w' = -3(1 + w)(c_s^2 - w). \tag{5.44}
\]

In a single fluid FRW model, with the dynamics fixed by an equation of state \(p = \omega \mu\) with \(\omega = \text{const.}\), \(c_s^2 = p/\mu\) is the speed of sound, and \(c_s^2 = \omega\). This is no longer the case when dealing with two or more fluids, where \(w' \neq 0\) and \(c_s^2(w) \neq \omega\) is not a proper speed of sound unless the fluids are tightly coupled [8].

In the case of one single fluid, the system is the same as that in section 5.3. In general, this set of equations (5.40), (5.43) and (5.44) forms a fourth order autonomous system, with (5.44) as an autonomous subsystem. As before, the order of the system can be reduced by the introduction of the variable \(\mathcal{U} = \Delta' / \Delta\) (5.29), resulting in the equation

\[
\mathcal{U}' = -\mathcal{U}^2 - \psi \mathcal{U} - \xi. \tag{5.45}
\]

A further reduction in the order of the system can be effected since, practically, either \(w\) or \(\Omega\) can be eliminated, reducing the phase space from \((\mathcal{U}, w, \Omega)\) to either \((\mathcal{U}, w)\) or \((\mathcal{U}, \Omega)\). Again, the equilibrium value of the variable \(\mathcal{U}\) determines whether the density inhomogeneity is increasing \(\mathcal{U} > 0\), stable \(\mathcal{U} = 0\) or decaying away \(\mathcal{U} < 0\).

\(^1\)There is a factor of 3 difference in some of the coefficients since Bruni and Piotrkowska define \(\tau\) as \(\tau = \ln a^3\).
Background Dynamics

It is possible to parametrize all the coefficients in the equations comprising the dynamical system in terms of \( w \), which allows a unified treatment of both flat and open models [8]. To achieve this, the scale factor is normalised at dust-radiation equidensity through the introduction of the variable \( S = a/a_E [43, 78] \).

Since the two fluids are only coupled through gravity, each component obeys the energy conservation equation separately. This gives \( \mu_d = \frac{\mu_E}{S^3} \) and \( \mu_r = \frac{\mu_E}{2S^4} \) for dust and radiation respectively, where \( \mu_E \) is the total energy density at equidensity. Total energy conservation gives

\[
\mu = \frac{\mu_E}{2} \left( \frac{1}{S^3} + \frac{1}{2S^4} \right). \tag{5.46}
\]

The total pressure derives from the radiation pressure only, giving \( p = p_r = \frac{\mu_r}{3} = \frac{\mu_E}{6S^4} \). This is sufficient to find \( w \) in terms of \( S \) [8]:

\[
w = \frac{1}{3(S+1)}, \tag{5.47}
\]

which can be inverted to give

\[
S = \frac{1 - 3w}{3w}. \tag{5.48}
\]

In this way the expansion of the universe model is parameterised by \( w \), with \( w \in [0, \frac{1}{3}] \) and varies from a pure radiation-dominated \((t \to 0)\) to a pure dust-dominated \((t \to \infty)\) phase.

As mentioned previously, for more than one fluid, \( c_s^2 = \dot{p}/\dot{\mu} \) is only formally the speed of sound. For dust and radiation, it is:

\[
c_s^2 = \frac{\dot{p}}{\dot{\mu}} = \frac{4}{3(4 + 3S)}, \tag{5.49}
\]

This gives the evolution of \( w \) (5.44) as:

\[
w' = w(3w - 1). \tag{5.50}
\]

At equidensity, \( S = 1 \), \( w = \frac{1}{6} \) and \( c_s^2 = \frac{4}{21} \). From (5.50) and (5.31), \( \Omega \) can be written in terms of \( w \) as:

\[
\Omega = \frac{3w\Omega_E}{3w\Omega_E + 2(1 - \Omega_E)(1 - 3w)^2}. \tag{5.51}
\]

67
This can be substituted into the expressions for \( \psi \) and \( \xi \), allowing these to be written in terms of \( w \) only. This reduces the phase space for the stability analysis to \( (\mathcal{U}, w) \) only, which is the phase space for the plane autonomous system (5.45), (5.50). The system (5.50), (5.43), which describes the evolution of the background universe model, is a plane autonomous system in the full space \( (\mathcal{U}, w, \Omega) \). This will always be a property of a dynamical system describing a perturbed universe, since the basic assumption of the perturbation analysis is that the back reaction of the perturbations on the dynamics of the background model is neglected [8].

The total energy density in terms of \( w \) is given by:

\[
\mu = \frac{27}{2} \mu E \frac{w^3}{(1-3w)^4},
\]

which also follows from the conservation equation \( \mu' = -3\mu(1+w) \), integrated using (5.50), with \( \mu \to \infty \) as \( w \to \frac{1}{3} \) and \( \mu \to 0 \) as \( w \to 0 \).

Using these results, and substituting in (5.49) for the dust-radiation background, the coefficients \( \psi (5.41) \) and \( \xi (5.42) \) in the differential equation (5.45) can be written as:

\[
\psi(w, \Omega) = \frac{1}{2} (1 + 3w)(1 - \Omega) + \frac{1 - 6w - 15w^2}{2(1 + w)},
\]

\[
\xi(w, \Omega) = -\frac{3}{2} \left[ 1 + \frac{w^2(5 - 3w)}{(1 + w)} \right] \Omega
+ \frac{4w(1 - 3w)}{(1 + w)} (1 - \Omega) + \Xi_K(w, k),
\]

where

\[
\Xi_K(w, k) = \frac{4w}{3(1 + w)} \frac{k^2}{a^2 H^2}
\]

has a different functional form depending on the curvature \( K \) [8]. In an open universe, \( a^2 H^2 = (1 - \Omega)^{-1} \) (see (5.7)), but in the flat case this equation cannot be used to make this substitution. Instead, (5.6) is used to make the substitution \( 3H^2 = \mu \), where \( \mu \) is given by (5.52). In this case, \( a = S \sigma E \) and \( S \) is given in (5.48). For the two different cases, the function \( \Xi \) is:

\[
K = 0, \quad \Xi = \frac{(1 - 3w)^2}{1 + w} \Xi_0 k^2, \quad \Xi_0 = \frac{8}{27a^2_E H^2_E},
\]

\[
K = -1, \quad \Xi = \frac{4w}{3(1 + w)} (1 - \Omega) k^2,
\]

\[
(5.56)
\]
Table 5.1: Equilibrium points for flat ($\Omega = 1$) models, with the corresponding eigenvalues

where $3H_E = \mu_E$ in the flat case, and $\Omega = \Omega(w)$ is substituted using (5.51) for open models. The definition (5.56) shows that, for $k \neq 0$, the function $\Xi_K$ is not continuous in the space $(\mathcal{U}, w, \Omega)$ on the $\Omega = 1$ plane, except on the line $w = \frac{1}{3}, \Omega = 1$. This discontinuity of $\Xi_K$ will play an important role in the behaviour of the perturbations [8].

### 5.4.1 Flat models

The flat models are described by the subsystem (5.45) and (5.50), substituting $\Omega = 1$ in $\psi$ (5.53) and $\xi$ (5.54), and using $\Xi$ for $K = 0$ given in (5.56). This definition shows that for flat models, a characteristic scale arises from the analysis that is related to the late time behaviour of the perturbation, i.e. the Hubble radius at equidensity, $H_E^{-1}$. For later convenience, we define

$$k_E = \frac{k}{a_E H_E} = 2\pi \frac{H_E^{-1}}{\lambda_E}, \quad (5.57)$$

which represents a wavenumber normalised at equidensity, since

$$k_E = 2\pi \Rightarrow \lambda_E = H_E^{-1}, \quad (5.58)$$

which corresponds to a perturbation wavelength which enters the horizon at the equivalence epoch [8].

Stationary points will exist on the $w = 0$ and $w = \frac{1}{3}$ axes, with

$$\mathcal{U}_\pm = \frac{1}{2} \left(-\psi_w \pm \sqrt{\psi_w^2 - 4\xi_w}\right), \quad (5.59)$$

69
where $\psi_w$, $\xi_w$ correspond to the stationary values of $\psi$ and $\xi$ for $w = 0, \frac{1}{3}$ [8]. The stationary points for the flat model are given in table 5.1.

As described in section 5.3.2, a solution having growing perturbations only occurs if there is a stable node on the positive side of the $\mathcal{U}$ axis. Equation (5.59) shows that in order to have $\mathcal{U}_+ \geq 0$, $\xi_0 \leq 0$, which implies

$$k_E \leq k_{EC}, \quad \text{where} \quad k_{EC} = \frac{9}{4}.$$  \hfill (5.60)

Also from (5.59), it is clear that stationary points only exist for $\psi_w^2 - 4\xi_w \geq 0$, which is always satisfied for $w = \frac{1}{3}$. For $w = 0$, the expression becomes

$$\psi_0^2 - 4\xi_0 = 6 \left( \frac{25}{24} - \frac{k_E^2}{k_{EC}^2} \right),$$  \hfill (5.61)

showing that stationary points only exist on this axis for perturbations with wavenumber $k_E \leq \frac{5}{2\sqrt{6}}k_{EC}$. This can be expressed as an equivalent wavelength restriction, $\lambda_E \geq \frac{2\sqrt{6}}{5}\lambda_{EC} \geq \lambda_{EC}$, where $\lambda_{EC} = \frac{8\pi}{9}H_E^{-1}$. This is the same condition for the existence of real eigenvalues, as can be seen from table 5.1. This follows from the fact that $w'$ does not depend on $\mathcal{U}$. Following the classification described in chapter 4, the values of the eigenvalues show that point 1 is an unstable node; point 2 is a saddle; point 3 is a saddle, and point 4 is a stable node [8]. The values of both the variable $\mathcal{U}$ and the

Figure 5.3: Phase space showing large scale perturbations in flat models. Here, as in the other figures $0 \leq w \leq \frac{1}{3}$, and $-4 \leq \mathcal{U} \leq 4$.
The eigenvalue $\lambda_U$ depend on the wavelength of the perturbation under consideration through the wavenumber $k_E$. Three different evolution regimes can be identified, i.e. $k_E < k_{EC}$, $k_{EC} < k_E < \frac{5}{2\sqrt{8}}k_{EC}$ and $k_E > \frac{5}{2\sqrt{8}}k_{EC}$, where $k_{EC}$ is the critical wavenumber.

The first of these, with $k_E$ smaller than the critical wavenumber $k_{EC}$ gives $U_+ > 0$. This corresponds to perturbations, with physical wavelength $\lambda_E > \lambda_{EC}$, which grow unbounded. This situation is shown in figure 5.3. In this figure, the saddle trajectories, which connect nodes with saddles, are clearly apparent. The trajectory ending in $U_-$ represents a purely decaying mode, whereas the other, ending in $U_+$, (the attractor), represents a purely growing mode. The generic trajectory represents a linear combination of the two modes, corresponding to the general solution of (5.45). Since $U$ is a tangent in the original phase space $(\Delta, \Delta')$, the trajectories in the figure which start from point 1 and go to the left of the saddle trajectory exit the figure on the left boundary, and re-enter it from the right, ending at point 4.

When point 4 does not exist, i.e. when $k_E > \frac{5}{2\sqrt{8}}k_{EC}$, the trajectories continue to enter and exit the figure many times. This characterizes the oscillatory behaviour of these modes in the $(U, w)$ plane. The perturbations represented by these trajectories oscillate as sound waves. Bruni and Piotrkowska [8] have checked that the amplitudes of these modes decay.
These are "small" perturbations, with physical wavelength \( \lambda_E < \frac{2\sqrt{6}}{5} \lambda_{EC} \). This situation is shown in figure 5.4.

Finally, there is an intermediate wavenumber regime with \( k_{EC} < k_E < \frac{5}{2\sqrt{6}} k_{EC} \). Here, point 4 has \( \mathcal{U}_+ < 0 \), which means that the stable node is located on the negative side of the \( \mathcal{U} \) axis. These solutions represent a decreasing inhomogeneity, where the perturbations in this range decay without oscillating, i.e. they are overdamped. The saddle trajectories are shown in figure 5.5, and represent the two solutions of (5.45) as before. The stability properties of dust-radiation flat models are a mixture of the properties of the pure dust and pure radiation models discussed in section 5.3, since perturbations of all sizes in a flat dust model grow unbounded, while all perturbations in a radiation model enter an oscillatory regime after they enter the sound horizon [8]. The appearance of the 3 different regimes, especially that of the intermediate, overdamped regime is a result of the 2-fluid approach.

5.4.2 Open models

In principle, these models are described by the full three-dimensional system given by equations (5.43), (5.45) and (5.50), with trajectories in the \((\mathcal{U}, w, \Omega)\) space. Given (5.51), we substitute for \( \Omega \) in the expressions for the coefficients \( \psi \) (5.53), \( \xi \) (5.54) and \( \Xi \) for \( K' = -1 \) (5.56). This reduces the analysis to
Table 5.2: Equilibrium points for open models with the corresponding eigenvalues

<table>
<thead>
<tr>
<th>Point 1: Unstable Node</th>
<th>Point 2: Saddle</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w = \frac{1}{3}, \quad \mathcal{U}_- = -1$</td>
<td>$w = \frac{1}{3}, \quad \mathcal{U}_+ = 2$</td>
</tr>
<tr>
<td>$\lambda_\mathcal{U} = 3$</td>
<td>$\lambda_\mathcal{U} = -3$</td>
</tr>
<tr>
<td>$\lambda_w = 1$</td>
<td>$\lambda_w = 1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Point 3: Saddle</th>
<th>Point 4: Stable Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w = 0, \quad \mathcal{U}_-0 = -1$</td>
<td>$w = 0, \quad \mathcal{U}_+0 = 0$</td>
</tr>
<tr>
<td>$\lambda_\mathcal{U} = 1$</td>
<td>$\lambda_\mathcal{U} = -1$</td>
</tr>
<tr>
<td>$\lambda_w = -1$</td>
<td>$\lambda_w = -1$</td>
</tr>
</tbody>
</table>

that of the two-dimensional system (5.45) and (5.50). Although this selects one particular open model in the class parametrized by $\Omega_E$, there is no loss of generality, since the dynamical properties of the models in the class do not depend on $\Omega_E$. For a given wavenumber $k$, the phase space evolution of all the models in the class is qualitatively the same. In terms of the geometry of the phase space $(\mathcal{U}, w, \Omega)$, the system (5.45),(5.50) corresponds to trajectories in the two-dimensional surface specified by $\Omega_E$ in this space [8]. We then consider the projection of this surface with its trajectories in the $(\mathcal{U}, w)$ plane. This is shown in figure 5.6:

Solutions to the system for open models are shown in table 5.2. These show that, contrary to the situation for flat models, the existence and nature of the stationary points for open models do not depend on the wavenumber $k$. All open models, therefore, share the same dynamical history. This results from the vanishing of the function $\Xi$ as $w \to 0$. This also implies that $\Omega \to 0$, since the trajectory is moving on the surface specified by $\Omega_E$. However, if (5.51) is not used to substitute for $\Omega$ in the three-dimensional system, then the limit $\Omega \to 0$ does not give the two-dimensional system for flat models. This is due to the discontinuity of the function $\Xi$ on the surface $\Omega = 1$, which was mentioned in section 5.4.

For the open model, point 4 is a stable node located at $\mathcal{U} = 0$. The system evolves towards this point since this is where the generic trajectory ends up, either directly from point 1, the unstable node, or by first exiting the left boundary of figure 5.6 and then re-entering from the right. Again, the saddle trajectory connecting points 1 and 3 represents the evolution of the purely decaying mode, whereas the trajectory from the saddle node at point 2 which ends at point 4 represents the purely growing mode. Since point 4 is located at $\mathcal{U} = 0$, all the perturbation modes evolve up to a
constant value, and then remain there. This is clear from the definition of $\mathcal{U}$ (5.29). Since the generic perturbation modes do not decay, open models are said to be stable, but not asymptotically stable [8]. This marginal stability against linear perturbations should not be taken to imply that structure cannot form in an open universe. It can be expected that this stability could be broken by non-linearity [8].

5.5 Perturbations of Bianchi Universes

The dynamical systems approach to density perturbations is not limited to FRW models, but can be applied to more general, anisotropic model. The simplest class of these is the class of Bianchi I solutions. Details of the analysis can be found in Dunsby [28]. As for the FRW models, the analysis begins with the exact equations (3.15) and (3.16). These are then linearized about a Bianchi I background. The simplest background to consider, is that for dust with zero vorticity, resulting in the restrictions

$$u_a = 0, \quad \omega_{ab} = 0, \quad w = 0, \quad c_s^2 = 0. \quad (5.62)$$

For a Bianchi I model, the shear does not vanish in the background spacetime ($\sigma_{ab} \neq 0$), and can therefore not be considered a first order variable. The linearized equations, therefore, are the same as the exact, perfect
The fluid equations would be (subject to the restrictions above) [112]:

\[
\dot{\Delta}_a = -Z_a - \sigma^c_a \Delta_c, \quad (5.63)
\]

\[
\dot{Z}_a = -2HZ_a - \sigma^b_a Z_b - \frac{1}{2} \mu \Delta_a - 2S_a, \quad (5.64)
\]

where

\[ S_a = a \tilde{\nabla} a \sigma^2. \quad (5.65) \]

These equations clearly contain terms which would be second order if linearizing about an isotropic background e.g. \( \sigma^c_a \Delta_c \). From these equations, a second order DE can be derived (see [112], & appendix C.3 for details). This can then be qualitatively analysed using the dynamical systems techniques described previously in this chapter. More details are given in [112]. In principle, the exact equations can be linearized about a large variety of backgrounds, making this approach quite general.

### 5.6 General applications of DS in Cosmology

In general relativity, the dynamical systems approach has been used mainly for studying cosmological models. When applied to the Einstein field equations, dynamical systems methods can give good qualitative information about the evolution of general classes of cosmological models, allowing for the exploration of a range of models, in order to see whether the simplest universe model is a probable solution to these equations, and giving insight into the full range of cosmological probabilities in epochs that are not constrained by observations [112]. In these sorts of investigations, the variables under consideration are the background variables themselves, e.g. variables describing the expansion, or curvature of the model under consideration.

In the study of Bianchi universes (see Appendix C.2.1), the use of dynamical systems methods, with expansion-normalized variables, leads to various insights. It highlights the importance of self-similar solutions, allows analysis of the dynamics of the models near the initial singularity, and, as part of the analysis, determines the stability properties of special solutions.

When regarded as evolution equations for inhomogeneous cosmologies, the Einstein field equations are partial differential equations, resulting in a state space which is an infinite dimensional function space. Many applications of DS in cosmological contexts involve the reduction of the evolution equations in an infinite dimensional space (partial differential equations) to evolution equations in a finite-dimensional subset (ordinary differential equations) [112].
For $G_2$ cosmologies (see Appendix C.2.2), this reduction process is performed by considering equilibrium states of the evolution equations in the infinite dimensional state space. In the perturbation analyses, described in this chapter, the reduction occurs because of the harmonic expansion of the spatial derivatives, allowing for a one-parameter family of differential equations, parametrized by the wave number $k$. 
Chapter 6

Large Scale Magnetic Fields

6.1 Introduction

Magnetic fields appear on all scales in our Universe, from the solar system, through interstellar and extra-galactic scales, to intra-cluster scales of several Mpc. Although magnetic field inhomogeneities have not yet been observed on scales as large as those exhibited by cosmic microwave background anisotropies, it is natural to expect that magnetic fields do indeed exist on such scales. Inflation provides the most natural explanation of field inhomogeneity on these scales, where the magnetic field may have an effect on structure formation [13]. Even if present magnetic fields do not have an important influence on large-scale structure, the possibility that they were important in the past still exists. Battaner and Florido [14] show that equivalent-to-present \(^1\) magnetic field strengths of the order of \(10^{-8} G\) could have played an important role in the physics of the evolution of structure formation.

Current research into galactic magnetic fields aims to answer questions such as how the observed large-scale organization of magnetic fields was achieved, to what extent and how do disk galaxies amplify their magnetic fields and what determines limiting field strengths achieved over a galaxy's lifetime. Whether the pre-galactic seed field strengths are fainter than \(10^{-11} G\) or stronger than \(10^{-8} G\) affects whether or not dynamo amplification is required to explain current observations. Important questions for structure formation include whether or not galaxies were the prime genera-

\(^1\)The strength which the magnetic field would have if it had evolved to the present epoch.
tors of intergalactic magnetic fields in the first place, or if significant fields were already generated in the pre-galactic era, and if the degree of ordering and strength of the interstellar magnetic field had a strong influence on the locations and dynamics of starforming regions within galaxies [63].

Much of our local universe consists of voids, regions containing very little baryonic matter. As yet, the weaker fields within these voids, which may be relics of true cosmic primordial fields, remain unmeasured. If these fields are indeed primordial relics, they could provide direct information about the very early universe. This is an exciting possibility, since the physical processes that took place in the very early universe do not have very many consequences that could still be directly detectable today. Most observables have been obliterated by the thermal bath of the pre-recombination era [49]. A primordial magnetic field with magnitude between $10^{-12}$ and $10^{-9} \text{G}$ could create significant magnetic fields in galaxy clusters through adiabatic compression alone, introducing new ingredients into the standard picture of the very early universe [99].

Looking at the possible dynamical significance of magnetic fields, some expectations seem reasonable. Large scale magnetic fields induce anisotropies into the expansion dynamics. Intergalactic fields will constrain the heat conductivity of the interstellar and intergalactic gas, and provide an additional component of pressure to it. Through magnetic fields, cosmic ray particles are coupled to the non-relativistic gas, and, depending on the magnetic energy density relative to the bulk dynamical and thermal energies of the gas, magnetic fields could strongly constrain the dynamics of galaxy formation [63].

### 6.2 Observations

Magnetic fields in intergalactic space are notoriously invisible, and can only be observed when "illuminated" by some form of detectable radiation. In the Milky Way and other late type galaxies, four ingredients allow for observational tracing of the large and small scale structure of the magnetic fields in these galaxies. These ingredients are (i) cosmic ray gas, (ii) magnetic fields, which generate polarized synchrotron radiation, (iii) ionized (and neutral) interstellar gas which causes Faraday rotation in the radio end of the spectrum, and (iv) interstellar grains which align with the interstellar magnetic field and induce linear polarization in the optical starlight by selective absorption or scattering [63].
6.2.1 Background

Awareness of the existence of magnetic fields beyond the solar corona began with the discovery, in 1949, of synchrotron radiation, an energy loss process for relativistic electrons in a magnetic field. Synchrotron emissivity does not give field strength unless combined with independent knowledge of the true number density of relativistic electrons. In the early 1950's the realization that the then recently-discovered non-thermal emission from the interstellar medium of the Milky Way was synchrotron radiation led to the conclusion that our galaxy possessed a magnetic field. Through the 1950's, with the development of radio astronomy, extended extragalactic radio sources (EGRS) with similar, non-thermal emission spectra were discovered. This implies that they too shine by synchrotron radiation, and hence must also possess an associated magnetic field [63]. In 1957 Bolton and Wild [23] suggested that the Zeeman splitting of a radio transition should be observable in the interstellar gas, providing a direct way of measuring the strength of this uniform magnetic field, given that

\[ \nu = \nu_{\text{in}} \pm eB(4\pi mc)^{-1}Hz, \]  

(6.1)

where \( \nu_{\text{in}} \) is a single transition, and \( B \) is in Gauss. That almost ten years passed between this suggestion and the first detections of this effect is an indication of the technical difficulty such measurements posed at the time [63]. Zeeman splitting has unfortunately not been observed in external galaxies or intergalactic space due to the Doppler smearing of the radio lines caused by dynamical effects, combined with the weak field strengths.

Faraday rotation is another detection method for magnetic fields. This is a measure of the rotation of the plane of polarization which occurs when plane polarized radiation (in this case in the radio) propagates through a plasma in which there is a component of the magnetic field parallel to the direction of propagation [17]. The rotation measure (RM) is given by [63]

\[ \frac{\Delta \chi}{(\Delta \lambda^2)} = 8.1 \times 10^5 \int n_e B_{||} dl \text{ rad.m}^{-2}, \]  

(6.2)

where \( \chi \) is the rotation (degrees) of the plane of polarization measured at wavelength \( \lambda(m) \), \( n_e \) (cm\(^{-3}\)) is the local density of non-relativistic electrons, \( B_{||} \) the line-of-sight component of magnetic field (G), and \( l \) the path length (pc). This equation shows that, in order to use Faraday RM to obtain field strengths, an independent measurement of the free electron density is required, as well as knowledge about its weighted distribution along the line.
of sight. Beginning in the 1960s, Faraday rotation measures from linear polarization of extragalactic radio sources led to the discovery that a large scale, organized magnetic field permeates the disk of our galaxy. Subsequent, more refined modelling of the large scale galactic magnetic field structure has been possible due to measurements of the rotation measures of ever larger numbers of extragalactic radio sources (see [63] and references therein). The availability of larger samples of extragalactic source rotations in the 1970s led to the first tests for Faraday rotation from a widespread cosmological magneto-ionic medium. As yet, no fields have been discovered, and the observations have established that any \textit{cosmologically aligned} magnetic field would have an upper strength limit of \( \approx 10^{-11} \) Gauss at the present epoch [63].

Over the last 20 years, observations have produced magnetic field detections not only in galaxy disks, but also in galaxy halos, clusters of galaxies, and in some distant galaxy systems which produce both absorption lines and Faraday rotation of the radiation from background quasars.

### 6.2.2 Current Observations

It is now known that a large scale, organized magnetic field fills the disk of the Milky Way. Studies of external galaxies indicate that all disc galaxies are permeated by large scale magnetic fields [113], and that \( \mu G \) level fields are common in spiral galaxy discs and halos. For the Milky Way, the average strength of the total field is \( 6 \pm 2 \mu G \) locally and \( 10 \pm 3 \mu G \) at 3 kpc Galactic radius. These values are derived from synchrotron data, using the assumption of equipartition of energy between cosmic rays and magnetic fields [24].

Intracluster magnetic fields can most easily be detected via synchrotron radiation of coextensive cosmic rays. This, like the Faraday RM, is not able to give a field strength estimate when used in isolation. Other probes of the intracluster medium are possible, including comparing the Faraday rotation of background radio sources shining through the cluster with comparable sources whose ray path avoids it. This allows testing for excess Faraday rotation due to the intracluster medium. This method was successfully used in 1990 [66] resulting in a measurement of \( 1.7 \pm 0.9 \mu G \) for the Coma cluster core region. Studies of a large number of galaxy clusters with many RM probes per cluster are currently unfeasible given the sensitivity limits of available radio telescopes, combined with the small angular size of more distant clusters. Obtaining RM probes through a sample of clusters, each
having typically only one or two (bright) polarized sources is a way of circumventing this problem. The consensus of these types of studies [26] is that cluster cores have a detectable component of RM, and many have field strengths at the microgauss level. Other measurements of the strengths of magnetic fields in intergalactic gas in 'normal' galaxy clusters have been obtained using Faraday RM combined with X-ray data. This gives typical field strengths for these fields of between 2 and 6μG, [64], which is comparable to the field strengths in the denser interstellar medium in our galaxy.

High resolution, multi-frequency VLA radio images of extended radio galaxies in the centres of dense cooling flow clusters have revealed extremely large Faraday rotations from kpc-scale ordered magnetic fields [63]. In 1993, Taylor and Perley [104] discovered that these field strengths can be as high as 30μG. These fields show ordered components on super-galactic scales [64]. These results seem to hint that, at least in cooling flow clusters, individual galaxies may form out of a strongly magnetized environment with prior magnetic field strength comparable to those found in the interstellar medium of galaxy disks.

As yet, Faraday rotation measures of high redshift objects only give upper limits (≈ 10^{-11}G in the present epoch) to fields on cosmological scales [63, 76]. The strength of any primordial magnetic field is also constrained by a number of astrophysical considerations. At the epoch of nucleosynthesis, a magnetized universe expands faster than a magnetism-free one due to the relativistic energy density supplied by the field. This leads to an increase in the synthesized abundance of helium-4, due to neutron-proton freeze-out of weak interactions occurring at a higher temperature [99]. Helium-4 observations can provide an upper limit on the energy density of any cosmological magnetic field at the time of nucleosynthesis. If the field is homogeneous over large scales, strong limits on the field strength at the time of last scattering are provided by the isotropy of the cosmic microwave background [3, 4, 5].

6.3 Origins of Cosmic Magnetic Fields

For a long time, a dynamo mechanism was the preferred explanation for the strengths of observed magnetic fields. The main alternative to the galactic dynamo is to assume that the galactic field results directly from a primordial field which is adiabatically compressed when the protogalactic cloud collapses [55]. Both of these mechanisms are essentially methods of amplifying a
pre-existing cosmological seed field whose required strength depends on the amplification method preferred.

6.3.1 Possible origins of seed field

There are at present over thirty theories about the origin of cosmic magnetic fields at galactic and intergalactic scales. Battaner & Lesch look at astrophysical arguments to examine these models [13], dividing them into four main categories, based on when the fields are generated, namely: a) during inflation, b) in a phase transition after inflation, c) during the radiation dominated era, and d) after recombination. It is the large scale fields which were produced during inflation which are most likely to have implications for structure formation.

Inflation

An inflation scenario for the creation of primordial magnetic fields was first proposed by Turner and Widrow [105]. The advantages of this scenario are [105, 55]:

- Inflation naturally produces effects on scales larger than the Hubble horizon, starting from causally connected microphysical processes. This could allow large-scale static magnetic fields to be generated by electromagnetic quantum fluctuations which have been amplified during inflation.

- The dynamical means to amplify these long wavelengths waves is also provided by inflation.

- Since the Universe is not a good conductor during inflation, (and perhaps during most of reheating) the magnetic flux is not conserved, and the ratio of the magnetic field to radiation energy density can increase.

- Classical fluctuations with wavelength $\lambda \leq H^{-1}$ of massless, minimally coupled fields can grow super-adiabatically i.e. their energy density decreases as $\approx a^{-2}$ instead of as $a^{-4}$.

The main problem with this scenario is that, in a conformally flat metric, such as FRW, the background gravitational field does not produce particles if the underlying theory is conformally invariant [55]. This is the case for photons, since classical electrodynamics is conformally invariant in the limit
of vanishing fermion masses. Turner and Widrow [105] discuss possible ways to break this conformal invariance, and to thereby get around this obstacle.

Phase transitions

Phase transitions of first order have been considered as potential mechanisms for the generation of primordial magnetic fields (see [13] and references therein). This phase transition would not take place simultaneously in all places in the Universe, but in causal bubbles. Among the phase transitions considered as candidates for generating magnetic fields are the electroweak phase transition and the quark-hadron phase transition (see [55]). Cosmic strings with wiggly motions formed in phase transitions may generate vorticity, and then magnetic fields. One of the major problems with phase transitions as magnetogenesis mechanisms is that they provide very small values of the magnetic field at galactic scales.

6.3.2 Amplification of a seed field

A dynamo is a mechanism for converting the kinetic energy of an electrically conducting fluid into magnetic energy. For a long time, the dynamo mechanism has been the conventional mechanism proposed for the amplification of small seed fields in galaxies. This mechanism occurs when the “frozen-in” term dominates the diffusion term in the equation describing the time evolution of the magnetic field, which is [63, 55]:

\[
\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) - \nabla \times \eta(\nabla \times \mathbf{B}),
\]

where \( \eta = \frac{1}{4\pi \sigma} \) is the magnetic diffusivity. If the field is frozen in, then \( \eta = 0 \), so that the second term vanishes, and the first term dominates, as is required for the dynamo mechanism.

The most common approach to the galactic dynamo uses the mean field approximation. This is necessary to accommodate the fact that the theory deals with averages over a range of correlation scales and times of the velocity field, and fluctuations in the magnetic field. It is based on the fact that fluctuations in the magnetic and velocity fields are much smaller than the mean of the slowly varying components of the corresponding quantities. The temporal evolution of the mean component of the magnetic field is given by [55]

\[
\frac{\partial B_0}{\partial t} = \nabla \times (\alpha B_0 + \mathbf{v}_0 \times B_0) - \nabla \times [(\eta + \beta)\nabla \times B_0],
\]

83
where the coefficient $\alpha$ is proportional to the helicity of the flow, and the term $\nabla \times (\beta \nabla \times B_0)$ describes the additional field dissipation due to turbulent motion. In equation (6.4), mean quantities are labelled by a subscript 0. The helicity (of outflowing magneto-plasma), through the $\alpha$-term, generates an electric field parallel to $B_0$. This field provides a mode for conversion of toroidal into poloidal magnetic field components called the $\alpha$-effect. In order to complete the dynamo cycle, $B_T \leftrightarrow B_P$, another mechanism is required to convert the poloidal field component into a toroidal one. This is provided by the differential rotation of the galactic disk, the $\omega$-effect, which, starting from a poloidal component, will wrap up the field line to produce a toroidal field. These two effects combined give rise to the so-called $\alpha - \omega$ galactic dynamo [55]. One of the main predictions of this mechanism is the generation of axially symmetric mean fields.

Doubts have been raised about the effectiveness of the diffusivity term in the standard dynamo theory. It has also been pointed out (see [63] and references) that the galactic disk field is tied largely to the densest clouds, which are ipso facto the most neutral clouds, hence the cloud field coupling is not strong enough to prevent the field from simply passing through the clouds with the most kinetic energy.

The main, although not the only, alternative to the galactic dynamo is to assume that the galactic field results from adiabatic compression of some primordial field during protogalactic cloud collapse [55]. Due to the large conductivity of the intergalactic medium, magnetic flux is conserved, implying that the magnetic field has to increase like the square of the size of the system. The required strength of the cosmic magnetic field at the galaxy formation time, adiabatically rescaled to present time is $\approx 10^{-16} G$, which is compatible with the observational limit on the field in the intergalactic medium derived by RMs, as well as with the big-bang nucleosynthesis constraints. This mechanism should wrap the field into a symmetric spiral structure with field reversal across the galactic disk diameter, and no reversal across the galactic plane [55].

### 6.4 Magnetic Fields and Structure Formation

Investigating the effect of magnetic fields on structure formation is not a recent endeavour. In 1971, Ruzmaikin & Ruzmaikina [84] gave a Newtonian analysis of the growth of density perturbations in a perfectly conducting medium with magnetic fields. Wasserman [107] considered magnetic influ-
enence on galaxy formation and angular momentum. Kim et al [65] extended this by including the back-reaction of the fluid on to the field, and as a result, derived a magnetic Jeans length. Battaner et al. [12] present relativistic analysis of the evolution of magnetic fields and their influence on density inhomogeneities in a radiation dominated universe, while Jedamzik et al. and Subramanian & Barrow [61] have considered magnetic dissipative effects at recombination.

More recently Tsagas and Barrow [97, 98, 99] have developed a covariant and gauge-invariant approach to the analysis of magnetized density perturbations, in a universe model containing a perfect fluid. Using the covariant gauge-invariant formalism of Ellis and Bruni [40], they derived a set of exact non-linear equations for general-relativistic magneto-hydrodynamic evolution. On linearizing these equations around a flat Friedman-Robertson-Walker (FRW) model, they were able to identify the general relativistic corrections to earlier Newtonian work, including the existence of a magnetic-curvature coupling.

Tsagas and Maartens [96] extended these results, considering also shape distortion effects due to the fields as well as density and rotational perturbations. They found explicit solutions to the perturbation equations for the radiation and dust eras, as well as pure-magnetic density perturbations. They also identified and analysed other sources of the magnetic effects.

The covariant analysis of density perturbations in a magnetized Bianchi I cosmology was undertaken as early as 1982 by Papadopulos and Esposito [77]. Their covariant approach was applied before the Ellis-Bruni gauge-invariant formalism was developed, and did not lead to quantitative results concerning the magnetic effects on structure formation. Tsagas and Maartens have used the covariant gauge-invariant approach in order to analyse perturbations on a magnetized Bianchi I background [100]. They found analytical solutions to the linearized perturbation equations and compared these to the FRW case to identify the corrections introduced by the background anisotropy. An investigation into the effects of magnetism on the expansion dynamics when magneto-curvature coupling is taken into account shows that this effect can reverse the pure magnetic effect on density perturbations [101].

In the next chapter, we present an investigation of magnetized cosmological perturbations in a flat FRW universe using a dynamical systems approach [59]. This allows us to investigate the overall dynamics and stability of the model with relative ease.
Chapter 7

A Dynamical Systems Approach to Magnetized Cosmological Perturbations

7.1 Overview of Chapter

This chapter is outlined as follows. In section 7.2, we use the GIC approach to perturbations outlined in chapter 4 to set up the evolution equations of the density and magnetic field inhomogeneities, and the curvature perturbations. We discuss in detail the approximations and linearization procedure in sections 7.3 & 7.4. In section 7.5, we set up a five dimensional autonomous dynamical system equivalent to the linearized propagation equations derived in section 7.2 - these equations give a detailed description of the full phase-space of solutions for perturbation dynamics in a magnetized dust-radiation universe. The analysis and discussion of this system, with particular emphasis on the three dimensional invariant sets representing the initial radiation and final dust dominated states are given in section 7.6.

7.2 Equations

7.2.1 Basic Equations

For all of the following analysis, we will use the covariant gauge invariant approach to perturbations presented in chapters 2 and 3. In particular, recall that the covariant derivative of any tensor may be split into derivatives along
and orthogonal to the fluid flow (see equations (2.9) & (2.10)), and that we can generalize the Newtonian operators of spatial divergence and curl to curved spacetime, as defined in equation (2.12).

The covariant derivative of the velocity 4-vector $u_a$ is split into its irreducible basic kinematic quantities i.e. the 4-acceleration, the expansion scalar, the shear tensor and the vorticity tensor (2.18). The evolution of the expansion scalar is described by Raychaudhuri’s equation (2.49), including a term for the magnetic energy density as follows:

$$\dot{\Theta} - A + 2(\sigma^2 - \omega^2) + \frac{1}{3} \Theta^2 + \frac{1}{2} (\mu + 3p + B^2) = 0,$$  \hspace{1cm} (7.1)

Local curvature is described by the Ricci tensor $R_{ab}$ while non-local tidal forces and gravitational radiation are described by the ‘electric’, $E_{ab}$ and ‘magnetic’, $H_{ab}$ parts of the Weyl tensor (2.41).

The electromagnetic field is represented, as usual, by the Maxwell tensor $F_{ab} = F_{[ab]}$ which splits into an electric part (2.61) and a magnetic part (2.62) when seen by an observer with 4-velocity $u_a$. Maxwell’s equations can be split into spatial and temporal parts, and written as:

$$(\text{div } E) - 2\omega^a B_a - q = 0,$$  \hspace{1cm} (7.2)

$$\dot{E}^{<a>} + j^a = \sigma^{ab} E_b + \epsilon^{abc} \omega_c E_b - \frac{2}{3} \Theta E^a + \epsilon^{abc} u_b B_c + (\text{curl } B)^a,$$  \hspace{1cm} (7.3)

$$(\text{div } B) + 2\omega^a E_a = 0.$$  \hspace{1cm} (7.4)

and

$$\dot{B}^{<a>} = \sigma^{ab} B_b - \epsilon^{abc} \omega_b B_c - \frac{2}{3} \Theta B^a - \epsilon^{abc} u_b E_c - (\text{curl } E)^a.$$  \hspace{1cm} (7.5)

### 7.2.2 Gauge Invariant Variables

In chapter 3, we defined two basic inhomogeneity variables: $\Delta_a = \frac{\delta \bar{\nabla}_a \mu}{\mu}$ (3.12) and $Z_a = \bar{\nabla}_a \Theta$ (3.14), the comoving spatial gradients of the energy density and the expansion, respectively. Following [96, 97], we define an additional variable to describe inhomogeneity in the magnetic field as follows:

$$B_a = \bar{\nabla}_a B^2,$$  \hspace{1cm} (7.6)
where $B^2 = B_a B^a$. This is the comoving spatial gradient of the field density $B^2$. Other basic variables which appear in the exact equations are the spatial gradient of the pressure, $Y_a = \nabla_a p$, the anisotropic pressure generated by the magnetic field, $\Pi_{ab} = \frac{1}{3} B^2 h_{ab} - B_a B_b$, and the comoving spatial gradient of the field vector: $B_{ab} = a \tilde{\nabla}_b B_a$.

7.2.3 Propagation and Conservation Equations

Description of the medium

The matter description under consideration is a mixture of dust and radiation interacting only through gravity. In general, the evolution equations for density perturbations couple to an entropy evolution equation through the equation of state $p = p(\mu, s)$, where $s$ is the entropy density (see equation (3.33)). However, since entropy perturbations are only important on very small scales, we will assume adiabatic perturbations at all times, thus ignoring any entropy contributions [30].

The total fluid equations take the same form as those for a perfect fluid of infinite conductivity. The infinite conductivity approximation allows the electric field to be omitted from Maxwell’s equations while spatial currents are preserved [97].

In this case, Maxwell’s equations (7.2) - (7.5) generate three constraints [96],

\[ \omega^a B_a = \frac{1}{2} q \]  (7.7)
\[ \text{curl} B^a = \epsilon^{abc} B_b \dot{u}_c + j^a, \]  (7.8)
\[ \text{div} B = 0, \]  (7.9)

and one propagation equation

\[ \dot{B}^{<a>} = \sigma^{ab} B_b + \epsilon^{abc} B_b \omega_c - \frac{2}{3} \Theta B^a. \]  (7.10)

Conservation Equations

Even though we are dealing with a two fluid medium, when considering only adiabatic perturbations, we only need the total fluid equations. Including the magnetic field energy density, the energy and momentum density conservation equations are respectively [97]:

\[ \dot{\mu} + 3\mu(1 + w)H = 0 \]  (7.11)
and
\[ \mu(1 + w + \frac{2B^2}{3\mu})\dot{u}_a + Y_a - \frac{2}{a} B_{[ab]}B^b + \dot{u}^b\Pi_{ba} = 0, \] (7.12)
where \( w = \frac{\dot{\rho}}{\rho} \). Ignoring entropy perturbations, the pressure can be written as a function of the energy density only, i.e. \( p = p(\rho) \). It follows that the relationship between \( Y_a \) and \( \Delta_a \) (3.33) becomes
\[ aY_a = \mu c_s^2 \Delta_a, \] (7.13)
where \( c_s^2 = \frac{d\rho}{d\mu} = \frac{\dot{\rho}}{\rho} \) is the speed of sound in the fluid.

**Exact propagation equations for the inhomogeneity variables**

The propagation equations for the various spatial gradients in the absence of a magnetic field were given in section 3.4. The following propagation equations, given in [97], include terms due to the presence of a magnetic field. They are, firstly, the equation for the *comoving fractional spatial gradient of the energy-density*:
\[
\dot{\Delta}_a = 3H\nu\Delta_a - (\sigma^a_a + \omega^b_a)\Delta_b - (1 + w)Z_a - 6H\frac{\mu}{\mu}B_{[ab]}B^b + \frac{2aHB^2}{\mu}u_a + \frac{a\Theta}{\mu}\dot{u}^b\Pi_{ba}, \tag{7.14}
\]
and secondly, the equation that governs the *spatial gradient of the expansion*:
\[
\dot{Z}_a = -2HZ_a - (\sigma^b_a + \omega^b_a)Z_b - \frac{1}{2}\mu\Delta_a + B_{ba}B^b + 3B_{[ab]}B^b + aR\dot{u}_a + \frac{3}{2}a\dot{u}^b\Pi_{ba} + aA_a - 2a\tilde{\nabla}_a(\sigma^2 - \omega^2), \tag{7.15}
\]
where \( A_a = h_a^b A_{b\dot{a}} = \tilde{\nabla}_a A, A = \tilde{\nabla}_a \dot{u}_a \) and
\[
\mathcal{R} = \frac{1}{2}K + A - 3(\sigma^2 - \omega^2), \tag{7.16}
\]
with
\[
K = \mu + \frac{1}{2}B^2 - \frac{1}{3}\Theta + \sigma^2 - \omega^2 \tag{7.17}
\]
representing the 3-Ricci scalar of the observers instantaneous rest space.
Finally, the equation describing the evolution of the orthogonal spatial gradient of the magnetic field is given by:

\[ a^{-2} \hat{h}^e_b h^p(a^2 B_{ep}) = -B_{ac}(\sigma^e_b + \omega^e_b) + \left( \sigma^e_a + \omega^e_a \right) B_{eb} \]

\[ - \frac{2}{3} B_a Z_b + 2a B_b \sigma^e_{[a} \dot{u}_{b]} \]

\[ - 2a B_c \omega^e_{(a} \dot{u}_{b)} + a B^e \nabla_b (\sigma_{ec} + \omega_{ac}) \]

\[ - H a (2 B_a \dot{u}_b + \dot{u}_a B_b) \]

\[ + a u^e B_a (\sigma_{ab} + \omega_{ab} + H h_{ab}) \]

\[ - a h_a \epsilon R_{epba} B^p u^q. \]  \hspace{1cm} (7.18)

### 7.3 Approximations

In the papers by Tsagas, Maartens and Barrow [96] - [99], the cosmic magnetic field is treated as a coherent test field propagating on the background. The large-scale magnetic field \( B_a \) is assumed to be too weak to destroy the spatial isotropy of the background spacetime, which is taken to be a flat Friedman-Robertson-Walker (FRW) model\(^1\). This is an acceptable physical approximation when the magnetic field density is small compared to the energy density of the fluid i.e. \( B^2 / \rho \ll 1 \). This issue is discussed in [114], where Zel'dovich calculates that \( B^2 / \rho < 8 \times 10^{-5} \) if the model is to be applicable during the whole of the radiation dominated era. Furthermore, the isotropy of the microwave background can be used to place strong limits on the magnitude of the magnetic field [4, 11].

Other authors have used alternative approaches to the problem of maintaining isotropy in the background in the presence of magnetic fields. Battaner et al. [13] show that a mean magnetic field is incompatible with a Robertson-Walker metric and therefore assume that there is no mean magnetic field on cosmological scales, i.e. \( < B_a > = 0 \). However, they include the presence of magnetic fields in smaller cells, with field directions random on larger scales. Thus, although there is no mean magnetic field, the model includes an average magnetic energy density \( < B^2 > \neq 0 \). Kim et al. [65] also follow this approach, considering that, at recombination, field directions are randomly oriented on scales smaller than the Hubble radius.

In order to maintain the coupling between magnetic irregularities and energy density perturbations in a straightforward way, a perfectly conducting medium is introduced. Looking at the covariant form of Ohm's law [97],

---

\(^{1}\)Spatial flatness is necessary for the 3-Ricci scalar to be gauge-invariant.
we have,

\[ J^a + J^b a^b = \sigma E^a, \quad (7.19) \]

where \( \sigma \) represents the conductivity of the medium. Projecting into the rest space of a fundamental observer yields

\[ J^{<a>} = \sigma E^a. \quad (7.20) \]

To have a vanishing electric field, while maintaining non-zero spatial currents \( (J^{<a>} \neq 0) \), the conductivity of the medium must be infinite, i.e. \( \sigma \rightarrow \infty \). In this infinite conductivity limit, the field is “frozen in” to the fluid. In this frozen-in condition, the fluid particles move with the magnetic field lines [115].

An alternative approach would be to assume a pure magnetic field with no electric field and no spatial currents. This would reduce (2.65) to

\[ \varepsilon^{abc} \dot{u}_b B_c + \text{curl} B^a = 0, \quad (7.21) \]

which is equivalent to

\[ \dot{u}_b B_c + \frac{1}{a} B_{[cb]} = 0. \quad (7.22) \]

Although this appears to couple the acceleration to the field gradient, when this relation is inserted into (7.12), all the magnetic terms cancel out. This effectively decouples the magnetic field from energy density inhomogeneities.

### 7.4 Linearization

#### 7.4.1 Linearization Scheme

As in section 3.5.3, we use an order-of-magnitude notation to linearize equations (7.14), (7.15) and (7.18). Because of the additional complication of the magnetic field, we introduce two smallness parameters [59]. The first, \( \epsilon_1 \) is the same as that used in section 3.5.3 to measure the extent to which the gauge-invariant variables deviate from zero (their value in a flat FRW universe). The other parameter \( \epsilon_2 \) is a measure of the Alfvén speed \( B^2/\mu \).

**Zero-order quantities**

The energy density \( \mu \), pressure \( p \) and expansion \( \Theta \) do not vanish in the background spacetime. The magnetic field, \( B_a \), is treated as a small test field propagating on the background. It follows that these variables can regarded as zeroth order in our approximation scheme.
First-order quantities [in $\epsilon_1$]

In order for the metric in some region of spacetime $\mathcal{U}$ to be written in a perturbed FRW form, the following inequalities must hold for the smallness parameter $\epsilon_1$ [99, 95]:

$$\frac{\sigma}{H} < \epsilon_1, \quad \frac{\omega}{H} < \epsilon_1, \quad \frac{|E_{ab}|}{H^2} < \epsilon_1, \quad \frac{|H_{ab}|}{H^2} < \epsilon_1,$$

$$\frac{|\tilde{\nabla}_{a\mu}|}{\mu H} < \epsilon_1, \quad \frac{|\tilde{\nabla}_a \Theta|}{H^2} < \epsilon_1.$$ (7.23)

where $|E_{ab}| \equiv (E_{ab} E^{ab})^{1/2}$, $|\tilde{\nabla}_a \rho| \equiv (\tilde{\nabla}_a \rho \tilde{\nabla}^a \rho)^{1/2}$ etc.

Tsagas [99] extends this definition to a magnetized universe by arguing that closeness to a spatially flat FRW spacetime is maintained when additional restrictions are imposed as follows:

$$\frac{|\Pi_{ab}|}{H^2} < \epsilon_1, \quad \frac{|\tilde{\nabla}_a B_a|}{H|B_a|} < \epsilon_1 \quad \text{and} \quad \frac{|K|}{H^2} < \epsilon_1.$$ (7.24)

In an exact flat FRW spacetime, all quantities of order $\epsilon_1$ vanish identically. Thus, $\sigma_{ab}$, $\omega_{ab}$, $\tilde{u}_a$, $\{A, A_a\}$, $\{E_{ab}, H_{ab}\}$, $\{\Delta_a, Z_a\}$, $\{B_{ab}, B_a\}$, $\{\kappa_a = a^2 K\}$ are all considered to be first order in $\epsilon_1$.

Note that although the magnetic field vector $B_a$ is considered to be a zeroth order quantity, its magnitude must remain small so that it does not disturb the isotropy of the background. This ensures that the the anisotropic pressure generated by the magnetic field, $\Pi_{ab}$, is negligible in the background, and therefore $\Pi_{ab}$ may also be regarded as first order in $\epsilon_1$.

Linearization of the above equations is implemented as follows: All terms higher than first order in $\epsilon_1$ are dropped, as well as terms higher than first order in the Alfvén parameter $\epsilon_2$. Terms like $\epsilon_1 \epsilon_2$ are however kept.

At the end of the calculation, terms first order in $\epsilon_2$ are dropped relative to zero order terms in the coefficients of quantities that are first order in $\epsilon_1$. This is permissible since the magnetic field is very weak ($B^2/\mu << 1$). However, this must be done last, otherwise terms may be dropped relative to others which could later vanish.
7.4.2 Linear Equations

Conservation equations

Linearization leaves equation (7.11) unchanged, but slightly modifies the momentum density equation (7.12):

\[ \mu (1 + w + \frac{2B^2}{3\mu}) \dot{u}_a + Y_a - \frac{2}{a} B_{[ab]} B^b = 0. \]  
(7.25)

This, together with (7.13), gives a useful expression for the acceleration vector:

\[ \dot{u}_a = \frac{1}{\left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)a} \left( \frac{2}{\mu} B_{[ab]} B^b - c_s^2 \Delta_a \right). \]  
(7.26)

It follows that the divergence of the acceleration \( A \) is given by \(^2\)

\[ A = \frac{1}{\left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} \left( \frac{c_s^2}{a^2} \Delta + \frac{B^2}{3\mu} K - \frac{B^2}{2\mu a^2} B \right). \]  
(7.27)

Propagation equations

The linearized propagation equations for the key gauge-invariant quantities are obtained by dropping terms of \( \mathcal{O}(2) \) or higher in both linearization parameters. Note that terms of mixed order \( (\epsilon_1 \epsilon_2) \) and first order in \( \epsilon_2 \) are kept to start with (but may be dropped later). To first order, all projected time derivatives of quantities that are first order in \( \epsilon_1 \) are equal to their normal time derivatives (e.g., \( \dot{\Delta}_{[a]} = \dot{\Delta}_a \)).

It follows that the energy density propagation equation (7.14) becomes:

\[ \dot{\Delta}_a = 3wH \Delta_a - (1 + w) \Theta_a - \frac{6H}{\mu} B_{[ab]} B^b + \frac{2aH B^2}{\mu} \dot{u}_a. \]  
(7.28)

The equation for the expansion gradient (7.15) involves the spatial gradient of the divergence of the acceleration \( A_a \). This can be determined using expression (7.27), the identities for commutations between spatial gradients and time derivatives given in appendix A, and the spatial gradient of the 3-curvature \( K \), given by

\[ a \bar{\nabla}_a K = 2\mu \Delta_a + aB_a - 4HZ_a. \]  
(7.29)

\(^2\)In deriving this, \( \bar{\nabla}_a w \) and \( \bar{\nabla}_a c_s^2 \) are treated as first order in \( \epsilon_1 \), as in [98].
Using these expressions, the evolution of the expansion gradient (to first order in $\epsilon_1$) is

$$\dot{\zeta}_a = -2Hz_a - \frac{1}{2}\mu\Delta_a - \frac{1}{2}aB_a - 3B_{(ab)}B^b$$

$$- \frac{c_s^2}{\left(1 + w + \frac{2B^2}{\mu}\right)} \tilde{\nabla}^2 \Delta_a - \frac{a}{2\mu \left(1 + w + \frac{2B^2}{\mu}\right)} \tilde{\nabla}^2 B_a$$

$$- \left[ \frac{6c_s^2(1 + w)}{\left(1 + w + \frac{2B^2}{\mu}\right)} + \frac{4B^2}{\mu \left(1 + w + \frac{2B^2}{\mu}\right)} \right] aH \tilde{\nabla}^b \omega_{ab}$$

$$+ \frac{2B^2}{3\mu \left(1 + w + \frac{2B^2}{\mu}\right)} \rho \Delta_a + \frac{B^2}{3\mu \left(1 + w + \frac{2B^2}{\mu}\right)} aB_a$$

$$- \frac{4B^2}{3\mu \left(1 + w + \frac{2B^2}{\mu}\right)} H \zeta_a .$$

(7.30)

### Scalar equations

Recall, from section 3.7, that the variable $\Delta_a$ contains more information than is necessary for the investigation of the growth or decay of density inhomogeneities. The local decomposition of $\Delta_a$ described in section 3.7 allows us to use the scalar variable $\Delta = a \tilde{\nabla}^a \Delta_a$ when investigating structure formation. We will also consider the following complementary scalar variables [98]:

$$\zeta = a \tilde{\nabla}^a Z_a, \quad B = \frac{a^2}{B^2} \tilde{\nabla}^a B^2, \quad K = a^2 K,$$

which represent spatial divergences in the expansion gradient, the energy density gradient of the magnetic field, and perturbations in the spatial curvature respectively.

As before, we linearize the scalar propagation equations by dropping terms of order $\epsilon_1^2$ or $c_s^2$, but retain all terms that are first order in $\epsilon_2$.

Equations describing the propagation of these scalars are:

$$\dot{\Delta} = 3wH \Delta - (1 + w)Z + \frac{3HB^2}{2\mu}B$$

$$- \frac{HB^2}{\mu} K - \frac{2c_s^2B^2}{\mu \left(1 + w + \frac{2B^2}{\mu}\right)} H \Delta ,$$

(7.32)
\[ \dot{z} = -2HZ - \frac{\mu}{2} \Delta - \frac{B^2}{2} Z + \frac{B^2}{4} \frac{c_s^2}{\left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} \nabla^2 \Delta \\
- \frac{B^2}{2\mu \left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} \nabla^2 B + \frac{2B^2}{3 \mu \left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} \mu \Delta \\
+ \frac{B^2}{3 \mu \left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} B^2 B - \frac{4}{3} \frac{B^2}{\mu \left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} HZ, \] (7.33)

\[ \dot{B} = -\frac{4}{3} Z + \frac{4HC_s^2}{\left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} \Delta - \frac{4HB^2}{3 \mu \left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} \mathcal{K} \\
+ \frac{2HB^2}{\mu \left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} B, \] (7.34)

and

\[ \dot{\mathcal{K}} = \frac{4HC_s^2}{\left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} \Delta + \frac{2HB^2}{\mu \left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} B \\
- \frac{4HB^2}{3 \mu \left(1 + w + \frac{2}{3} \frac{B^2}{\mu}\right)} \mathcal{K}. \] (7.35)

### 7.4.3 Final Linearized System

Combining equations (7.32 - 7.35), and finally dropping terms of order \( \epsilon_2 \) with respect to zero order quantities, we obtain a second-order differential equation for the scalar energy density perturbations:

\[ \ddot{\Delta} = -(2 + 3c_s^2 - 6w)H \dot{\Delta} + \frac{1}{2} \left(1 - 6c_s^2 + 8w - 3w^2\right) \mu \Delta \\
+ c_s^2 \nabla^2 \Delta - \frac{2B^2 H}{\mu \left(1 + w\right)} c_s^2 \Delta - \frac{1}{2} \left(1 - 3c_s^2 + 2w\right) \mu c_s^2 B \\
+ \frac{1}{2} c_s^2 \nabla^2 B + \frac{1}{3} \left(2 - 3c_s^2 + 3w\right) \mu c_s^2 \mathcal{K}. \] (7.36)

It is not immediately obvious that the term containing \( c_s^2 \) is negligible with respect to the other terms involving \( \Delta \). In order to clarify this, we can write
\( \dot{c}_s^2 \) in terms of \( w \) by noting

\[
\dot{c}_s^2 = \frac{d(c_s^2)}{dw} w = \frac{4(w - c_s^2)}{(1 + w)} H .
\]  

(7.37)

In what follows we will consider the above equations for values of \( w \in [0, \frac{1}{3}] \), but more specifically, we will look at the dust \( w = 0 \) and radiation dominated \( w = \frac{1}{3} \) eras. From equation (7.50) we see that \( w = 0 \Rightarrow c_s^2 = 0 \) and that \( w = \frac{1}{3} \Rightarrow c_s^2 = \frac{1}{3} \). Thus, the term drops away during both these eras and is negligible at all other times.

Without that term, the second order propagation equation for \( \Delta \) becomes

\[
\ddot{\Delta} = -(2 + 3c_s^2 - 6w)H \dot{\Delta} + \frac{1}{2}(1 - 6c_s^2 + 8w - 3w^2) \mu \Delta \\
+ c_s^2 \nabla^2 \Delta - \frac{1}{3}(1 - 3c_s^2 + 2w) \mu \alpha (\alpha + \frac{1}{2}c_s^2) \nabla^2 B \\
+ \frac{1}{3}(2 - 3c_s^2 + 3w) \mu \alpha (\alpha - c_s^2) \nabla^2 \mathcal{K} .
\]  

(7.38)

Finally, the completely linearized propagation equations for \( B \) and \( \mathcal{K} \) are [59]:

\[
\dot{B} = \frac{4}{3(1 + w)} \dot{\Delta} + 4H \frac{(c_s^2 - w)}{(1 + w)} \Delta ,
\]  

(7.39)

and

\[
\dot{\mathcal{K}} = \frac{4Hc_s^2}{(1 + w)} \mathcal{K} - \frac{4}{3 \mu (1 + w)} B^2 H \mathcal{K} + \frac{2B^2 H}{\mu (1 + w)} B .
\]  

(7.40)

At this point Tsagas and Maartens [96] expand the LHS of equation (7.40) using the definition of \( \mathcal{K} \) (see equation (7.31)): \( \dot{\mathcal{K}} = 2H \mathcal{K} + a^2 \dot{K} \) and then drop the second term on the RHS of (7.40) relative to \( 2HK \). Due to the slightly different linearization adopted here, equation (7.40) has an extra term compared to the corresponding result in [96]. This term was identified in [98], and the role of the magnetic tension as the physical reason behind it has been discussed in [100]. Equations (7.38) and (7.39) are identical to those in [96].

### 7.5 The Dynamical System

As in chapter 5, we change the independent variable from proper time to the function \( \tau \) of the scale factor \( a \) defined in equation (5.2). This yields
\( \frac{d\tau}{dt} = H \). Using the harmonic decomposition described in section 3.9, equations (7.38), (7.39) and (7.40) become

\[
\Delta'' = -\frac{1}{2}(1 + 6c_s^2 - 15w)\Delta' + \frac{3}{2}(1 - 6c_s^2 + 8w - 3w^2)\Delta \\
- c_s^2 \frac{k^2}{a^2 H^2} \Delta - \frac{3}{2}(1 - 3c_s^2 + 2w)c_s^2 B - \frac{c_s^2}{2} \frac{k^2}{a^2 H^2} B \\
+ (2 - 3c_s^2 + 3w)c_s^2 \mathcal{K},
\]

(7.41)

\[
B' = \frac{4}{3(1 + w)} \Delta' + \frac{4(c_s^2 - w)}{(1 + w)} \Delta,
\]

(7.42)

and

\[
\mathcal{K}' = \frac{4c_s^2}{(1 + w)} \Delta - \frac{4c_s^2}{3(1 + w)} \mathcal{K} + \frac{2c_s^2}{(1 + w)} B,
\]

(7.43)

where a prime denotes differentiation with respect to \( \tau \).

### 7.5.1 Coefficients in terms of \( w \)

In order to close the system, propagation equations are needed for all background variables. It turns out however that the coefficients of the above differential equations can all be written explicitly as functions of the equation of state parameter \( w \), thus only an evolution equation for \( w \) needs to be found. To achieve this, we normalize the scale factor \( a \) at dust-radiation equi-density by using the variable \( S = a/a_F \) that was introduced in section 5.4 (see also Ehlers & Rindler [43] and Padmanabhan [78]).

Since the two fluids are coupled only through gravity, the energy conservation equation is thus obeyed separately for each component, making it straightforward to find \( w \) in terms of \( S \) as before (see section 5.4):

\[
w = \frac{1}{3(S + 1)},
\]

(7.44)

which can be inverted to give

\[
S = \frac{1 - 3w}{3w}.
\]

(7.45)

In this way the expansion of the universe model is parameterised by \( w \), with \( w \in [0, \frac{1}{3}] \) and varies from a pure radiation-dominated (\( t \to 0 \)) to a pure dust-dominated (\( t \to \infty \)) phase.
These results (7.44) and (7.45) allow us to write the coefficients that occur in the perturbation equations as functions of \( w \) only. Recall that \( c_s^2 \) is given by

\[
c_s^2 = \frac{4w}{3(1 + w)}.
\] (7.46)

The energy density of the magnetic field \( B^2 \) has a radiation-like propagation equation \( B^2 = B_E^2 S^{-4} \), hence the Alfvén speed can be written as

\[
c_a^2 = 6w a_B^2.
\] (7.47)

and the total energy density is given by

\[
\mu = \frac{27}{2} \mu_E \frac{w^3}{(1 - 3w)^3}.
\] (7.48)

This leads to a similar equation for the square of the Hubble parameter \( H^2 \):

\[
H^2 = \frac{27}{2} H_E^2 \frac{w^3}{(1 - 3w)^3},
\] (7.49)

where \( H_E \) is the value of the Hubble parameter at matter-radiation equality. The propagation equation for \( w \) can now be written down. It is

\[
w' = 3w(w - \frac{1}{3}).
\] (7.50)

When the perturbation equations are harmonically decomposed, the Laplacian terms give rise to coefficients of the form \( k^2 / a^2 H^2 \). These can be expressed in terms of the variable \( w \) as follows:

\[
\frac{k^2}{a^2 H^2} = \frac{k^2}{a_E^2 H_E^2} \frac{2(1 - 3w)^2}{3w}
\]

\[
= \frac{k_E^2}{a_E^2 H_E^2} \frac{2(1 - 3w)^2}{3w},
\] (7.51)

where \( k_E^2 = \frac{k^2}{a_E^2 H_E^2} \).

In terms of \( w \), (7.41) becomes

\[
\Delta'' = \alpha \Delta' + \beta \Delta + \gamma B + \eta \mathcal{K},
\] (7.52)
\[
\alpha_\Delta = -2\left(\frac{1-3w^2}{1+w}\right) + \frac{3}{2}(1+w),
\]
\[
\beta_\Delta = \frac{3}{2} \left[ 1 + \frac{w(5-3w)}{1+w} \right] - \frac{8(1-3w)}{9(1+w)} k_E^2,
\]
\[
\gamma_\Delta = -9w c_a E \left[ \frac{1-w+2w^2}{1+w} \right] - 2c_a^2 E (1-3w)k_E^2,
\]
\[
\eta_\Delta = 6w c_a^2 E \left[ \frac{2+w+3w^2}{1+w} \right].
\]
(7.53)

Similarly, (7.42) and (7.43) become
\[
B' = \alpha_B \Delta' + \beta_B \Delta,
\]
(7.54)

and
\[
\mathcal{K}' = \alpha_K \Delta + \beta_K \mathcal{K} + \gamma_K B,
\]
(7.55)

where
\[
\alpha_B = \frac{4}{3(1+w)},
\]
\[
\beta_B = \frac{4w}{3(1+w)^2} (1-3w),
\]
\[
\alpha_K = \frac{16w}{3(1+w)^2},
\]
\[
\beta_K = \frac{8w}{(1+w)c_a E},
\]
\[
\gamma_K = \frac{12w}{(1+w)c_a^2 E},
\]
(7.56)

with the evolution of the background determined by (7.50).

### 7.5.2 Final system

In the simplest case of a single fluid, equations (7.52), (7.54) and (7.55) together with the condition \(w = \text{constant}\), form a fourth-order autonomous system of differential equations. In the more general two-component case, the system is five dimensional due to the inclusion of the propagation equation for \(w\) (7.50). However, the order of the system can be reduced by using
the variable \( \mathcal{U} \), defined in equation (5.29) as

\[
\mathcal{U} = \frac{\Delta'}{\Delta}.
\]  

(7.57)

In order to keep the system dimensionally consistent, we define two new variables,

\[
V = \frac{B}{\Delta}, \quad \text{and} \quad W = \frac{K}{\Delta}.
\]  

(7.58)

In terms of \( \mathcal{U}, V, W \), the equations (7.52), (7.54) and (7.55) become

\[
\mathcal{U}' = -\mathcal{U}^2 + \alpha_\Delta \mathcal{U} + \beta_\Delta + \gamma_\Delta V + \eta_\Delta W,
\]  

(7.59)

\[
V' = \alpha_B \mathcal{U} + \beta_B - VU,
\]  

(7.60)

\[
W' = \alpha_K + \beta_K W + \gamma_K V - WU,
\]  

(7.61)

with the coefficients determined by (7.53), & (7.56), together with the propagation equation for \( w \) (7.50).

### 7.6 Analysis

The system is non-linear, however we can analyse it locally by linearizing about any stationary points, without losing the details of the qualitative dynamics [75, 112].

From equation (7.50) we can see that stationary points exist for \( w = \frac{1}{3} \) and \( w = 0 \). Since equation (7.50) decouples from the rest of the system, the 3-dimensional subsystems corresponding to these values of \( w \) correspond to invariant sets describing the early radiation-dominated and late dust-dominated periods of dynamical evolution. We now consider these cases separately.

#### 7.6.1 Radiation era: \( w = \frac{1}{3} \)

On substituting \( w = \frac{1}{3} \) into (7.53) and (7.56), the system for the radiation-dominated era becomes:

\[
\mathcal{U}' = -\mathcal{U}^2 + \mathcal{U} + 2 - 2c_{ab}^2 V + 4c_{ab}^2 W,
\]  

\[
V' = \mathcal{U} - VU,
\]  

\[
W' = 1 - 2c_{ab}^2 W + 3c_{ab}^2 V - WU.
\]  

(7.62)
<table>
<thead>
<tr>
<th>Point</th>
<th>Values ((w, U, V, W))</th>
<th>Eigenvalues ((\lambda_{1,2,3,4}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>(\left(\frac{1}{3}, 0, -\frac{1}{c_{aE}}, -\frac{1}{c_{aE}}\right))</td>
<td>(\frac{1}{3} - 4c_{aE}^2) 2 (-1 + 2c_{aE}^2)</td>
</tr>
<tr>
<td>R2</td>
<td>(\left(\frac{1}{3}, -4c_{aE}^2, 1, -\frac{1}{2} (c_{aE}^2 + 3)\right))</td>
<td>(\frac{1}{3} 4c_{aE}^2) 2 + 8c_{aE}^2 (-1 + 2c_{aE}^2)</td>
</tr>
<tr>
<td>R3</td>
<td>(\left(\frac{1}{3}, -1 + 2c_{aE}^2, 1, -(1 + 7c_{aE}^2)\right))</td>
<td>(\frac{1}{3} 1 - 2c_{aE}^2) 1 - 6c_{aE}^2 (3 - 2c_{aE}^2)</td>
</tr>
<tr>
<td>R4</td>
<td>(\left(\frac{1}{3}, 2, 1, \frac{1}{2} (1 + 2c_{aE}^2)\right))</td>
<td>(\frac{1}{3} -2) (-2 - 4c_{aE}^2) (-3 + 2c_{aE}^2)</td>
</tr>
</tbody>
</table>

Table 7.1: Table of equilibrium points for \(w = \frac{1}{3}\)

The equilibrium points for (7.62) can easily be determined and are given in table 7.1 above.

The 3-dimensional space described by \(\{U, V, W\}\), and \(w = \frac{1}{3}\) (see figure I) is an invariant set, so we can classify the equilibrium points according to the eigenvalues \(\lambda_{2,3,4}\) (since \(\lambda_1 = \frac{1}{3}\) is not relevant).

**Point R1** is a saddle (not included in figure I because it is far away from the other points). Orbits close to it may initially evolve towards it, but end up evolving away again. At R1 \(U = 0\), so density perturbations neither grow or decay. Using equations (7.57) and (7.58) it follows that \(\Delta\), \(B\) and \(\mathcal{K}\) are constant:

\[
\begin{align*}
\Delta &= \text{const} = C_0, \\
B &= -\frac{1}{c_{aE}} C_0, \\
\mathcal{K} &= -\frac{1}{c_{aE}} C_0.
\end{align*}
\]

(7.63)

**Point R2** is also a saddle, but this time since \(U < 0\), density inhomogeneities are decreasing. Solutions at R2 again follow from equations (7.57) and (7.58):

\[
\begin{align*}
\Delta &= C_1 \left(\frac{a}{a_{E}}\right)^{-4c_{aE}^2}, \\
B &= \Delta = C_1 \left(\frac{a}{a_{E}}\right)^{-4c_{aE}^2}, \\
\mathcal{K} &\approx -\frac{1}{2c_{aE}^2} C_1 \left(\frac{a}{a_{E}}\right)^{-4c_{aE}^2}.
\end{align*}
\]

(7.64)
Point R3 is a node source, representing unstable equilibrium, so all orbits close to this point evolve away from it. Again since $\mathcal{U} = -1 + 2c_{\alpha E}^2 < 0$, solutions at R3 represent a decreasing density inhomogeneity:

$$
\Delta = C_2 \left( \frac{a}{a_E} \right)^{(-1+2c_{\alpha E}^2)}, \\
B = \Delta = C_2 \left( \frac{a}{a_E} \right)^{(-1+2c_{\alpha E}^2)}, \\
K = -(1 + 7c_{\alpha E}^2)C_2 \left( \frac{a}{a_E} \right)^{(-1+2c_{\alpha E}^2)}.
$$

(7.65)

Point R4 is a stable node, or a sink. Orbits close to this point evolve towards it as $\left( \frac{a}{a_E} \right)$ increases. Since $\mathcal{U} > 0$ solutions at R4 represent growing density inhomogeneities:

$$
\Delta = C_3 \left( \frac{a}{a_E} \right)^2, \\
B = \Delta = C_3 \left( \frac{a}{a_E} \right)^2, \\
K = \frac{1}{2}(1 + 2c_{\alpha E}^2)C_3 \left( \frac{a}{a_E} \right)^2.
$$

(7.66)

Solutions at the points given above correspond to approximate solutions (to leading order in $c_{\alpha E}^2$) of the perturbation equations during the radiation-dominated era. Therefore by linearity, the general solutions for the perturbation variables $\Delta$, $B$ and $K$ are given by a linear combination of (7.63 - 7.66):

$$
\Delta = C_0 + C_1 \left( \frac{a}{a_E} \right)^{-4c_{\alpha E}^2} + C_2 \left( \frac{a}{a_E} \right)^{(-1+2c_{\alpha E}^2)} + C_3 \left( \frac{a}{a_E} \right)^2 \\
B = -\frac{1}{c_{\alpha E}^2}C_0 + C_1 \left( \frac{a}{a_E} \right)^{-4c_{\alpha E}^2} + C_2 \left( \frac{a}{a_E} \right)^{(-1+2c_{\alpha E}^2)} + C_3 \left( \frac{a}{a_E} \right)^2, \\
K \approx -\frac{1}{c_{\alpha E}^2}C_0 - \frac{1}{2c_{\alpha E}^2} \left( \frac{a}{a_E} \right)^{-4c_{\alpha E}^2} \\
- (1 + 7c_{\alpha E}^2)C_2 \left( \frac{a}{a_E} \right)^{(-1+2c_{\alpha E}^2)} \\
+ \frac{1}{2}(1 + 2c_{\alpha E}^2)C_3 \left( \frac{a}{a_E} \right)^2.
$$

(7.67)
Table 7.2: Table of equilibrium points for $w = 0$, with $\xi = \frac{1}{2} \sqrt{6 (\frac{25}{24} - \frac{k_B^2}{l_{EC}^4})}$.

This can be written more concisely as:

\[
\Delta = C(0) + \sum_\alpha C(\alpha) (\frac{a_E}{a})^\alpha, \\
B = -\frac{1}{a_E} C_0 + \sum_\alpha C(\alpha) (\frac{a}{a_E})^\alpha, \\
K = -\frac{1}{a_E^2} + (1 + 3c_{aE}^2) \sum_\alpha \frac{C(\alpha)}{(\alpha + 2c_{aE}^2)(\frac{a}{a_E})^\alpha}, \tag{7.68}
\]

where $\alpha$ solves the cubic equation:

\[
\alpha^3 + (2c_{aE}^2 - 1)\alpha^2 - 2\alpha - 8c_{aE}^2(1 + c_{aE}^2) = 0, \tag{7.69}
\]

and corresponds to the super-horizon solutions given in [96], with slightly modified exponents - due to the extra term in the equation for the spatial curvature $K$ (7.43).
7.6.2 Dust era: \( w = 0 \)

When solving for the equilibrium points in the dust-dominated era, we substitute \( w = 0 \) into the equations defining the dynamical system (7.59-7.61). This yields:

\[
\begin{align*}
\mathcal{U}' & = -\mathcal{U}^2 - \frac{1}{2} \mathcal{U} + \frac{3}{2} - \frac{8}{9} k_E^2 - 2c_{oE}^2 k_E^2 V , \\
V' & = \frac{4}{3} \mathcal{U} - V \mathcal{U} , \\
W' & = -W \mathcal{U} .
\end{align*}
\]  

(7.70)

The equilibrium points of this system are shown in table 7.2. In this case, the \( \mathcal{U} \) and \( V \) propagation equations decouple from the equation for \( W' \), which means that the system is effectively only 2-dimensional. A critical scale \( \lambda_{EC} \equiv 2 \pi a / k_{EC} \) appears in the analysis of this equation through its corresponding wave number \( k_{EC} \):

\[
k_{EC}^2 = \frac{27}{16} \frac{1}{1 + 3c_{oE}^2} \approx \frac{27}{16} (1 - 3c_{oE}) .
\]  

(7.71)

The solutions at each point will have different behaviour depending on the value of \( k_E \). In what follows, we first look at the general properties of the equilibrium points and their corresponding solutions and then give specific information about their behaviour for the three regimes of \( k_E \) values that emerge from the analysis.

**Point D1:** These solutions are confined to the \( \{\mathcal{U}, V\} \) plane for constant \( W \) and approach a constant value independent of the value of \( k_E \):

\[
\begin{align*}
\Delta & = \text{const} = C_0 , \\
B & = -\frac{1}{2c_{oE}^2} \left[ \frac{3}{2} \right] C_0 , \\
\mathcal{K} & = \text{const} = C_k .
\end{align*}
\]  

(7.72)

**Point D2:** This point has different stability behaviour for different values of \( k_E \). The curvature variable \( W \) vanishes independently of \( k_E \). Since the solution at the point depends on \( \xi = \xi(k_E) \), the nature of the solution mode (e.g. growing/decaying/oscillation) will depend on the value of \( k_E \):

\[
\begin{align*}
\Delta & = C_1 (\frac{a}{a_E})^{(\frac{1}{4} + \xi)} , \\
B & = \frac{4}{3} C_1 (\frac{a}{a_E})^{(-\frac{1}{4} + \xi)} , \\
\mathcal{K} & = 0 .
\end{align*}
\]  

(7.73)
Point D3: This point has different stability behaviour in two different regions of $k_E$. In both regions, the curvature variable vanishes. The solutions at D3 are:

$$\Delta = C_2\left(\frac{a}{a_E}\right)^{-\frac{1}{4}} - \xi',
\quad B = \frac{4}{3} C_2\left(\frac{a}{a_E}\right)^{-\frac{1}{4}} - \xi,
\quad \mathcal{K} = 0. \quad (7.74)$$

As in the radiation dominated era (discussed in section 7.6.1), solutions (to leading order in $\frac{C_2}{a_E}$) for the perturbation variables $\Delta$, $B$ and $\mathcal{K}$ are given by a linear combination of the solutions at D1, D2 and D3:

$$\Delta = C_0 + C_1\left[\frac{2}{3k_E^2} - \frac{8}{3}\right] C_0 + \frac{4}{3} C_1\left(\frac{a}{a_E}\right)^{-\frac{1}{4}} + \frac{4}{3} C_2\left(\frac{a}{a_E}\right)^{-\frac{1}{4}} - \xi,
\quad B = \frac{1}{2a_E}\left[\frac{2}{3k_E^2} - \frac{8}{3}\right] C_0 + \frac{4}{3} C_1\left(\frac{a}{a_E}\right)^{-\frac{1}{4}} + \frac{4}{3} C_2\left(\frac{a}{a_E}\right)^{-\frac{1}{4}} - \xi,
\quad \mathcal{K} = \text{const} = C_k. \quad (7.75)$$

We now look more closely at how the nature of the stationary points D1, D2 and D3 and their corresponding solutions depends on the wavenumber $k_E$.

Region 1 - $k_E \leq k_{EC}$. See figure 7.2

Point D1: In this region, the eigenvalues are, respectively, $\lambda_2 = 0$, $\lambda_3 \geq 0$, and $\lambda_4 < 0$. It follows that D1 corresponds to a line of saddles in the solution space for $\{U, V, W\}$.

Point D2: Here, all the eigenvalues are either negative or zero, so this point is a stable sink. Since $\xi \geq \frac{1}{4} \Rightarrow U \geq 0$, it follows that the solution at D2 corresponds to a growing density inhomogeneity.

Point D3: All eigenvalues are positive, so this point is an unstable node. Since $U = -\frac{1}{4} - \xi < 0$, it follows that the solution at D3 corresponds to a decaying density inhomogeneity. For very long wavelength solutions i.e. $k_E << k_{EC}$ we have $\xi \approx \frac{3}{4}$. In this case the explicit solutions of the perturbation equations (which are a linear combination of the solutions at the points D1, D2 and D3) for the dust era are given by:
Figure 7.2: Dust dominated era - $k_E \leq k_{EC}$.

\[
\Delta = C_0 + C_1 \left( \frac{a}{a_E} \right) + C_2 \left( \frac{a}{a_E} \right)^{-\frac{3}{2}}
\]

\[
B = \frac{1}{2} \left[ \frac{\Xi}{2k_E^2} - \frac{3}{8} \right] C_0 + \frac{4}{3} C_1 \left( \frac{a}{a_E} \right) + \frac{4}{3} C_2 \left( \frac{a}{a_E} \right)^{-\frac{3}{2}}
\]

\[
\mathcal{K} = C_k.
\]  

(7.76)

In the dust era, \( \left( \frac{a}{a_E} \right) = \left( \frac{t}{t_E} \right)^{2/3} \), so we can write the above solutions (7.76) in terms of \( \left( \frac{t}{t_E} \right) \) as in [96]. For example, the solution for density perturbations is:

\[
\Delta = C_0 + C_1 \left( \frac{t}{t_E} \right)^{2/3} + C_2 \left( \frac{t}{t_E} \right)^{-1}.
\]

It consists of a magnetic field induced constant mode, plus the two usual non-magnetized adiabatic modes. Comparing to [96] equation (65), we see that the dynamical systems analysis does not recover the non-adiabatic decaying mode. This is because our analysis is set up to look for asymptotic solutions, so as $t \to \infty$, the magnetic energy density decays faster than that of the fluid, and $c_a^2 \to 0$. This removes the magneto-curvature coupling from the propagation equation for $\mathcal{K}$, and therefore our analysis does not obtain the extra decaying mode induced by the magnetic field.

106
Figure 7.3: Dust dominated era - $k_{EC} < k_E \leq 5/2\sqrt{6}k_{EC}$.

Region 2 - $k_{EC} < k_E \leq \frac{5}{2\sqrt{6}}k_{EC}$. See figure 7.3

**Point D1:** In this region, the eigenvalues are, respectively, $\lambda_2 = 0$, $\lambda_3 < 0$ and $\lambda_4 < 0$. It follows that D1 represents a line of sinks (only one is shown in figure 7.3) in the solution space for $\{U, V, W\}$.

**Point D2:** In this region, the eigenvalues are, respectively $\lambda_2 \leq 0$, $\lambda_3 > 0$ and $\lambda_4 > 0$, so D2 is a saddle point (or a line of sources if $\lambda_2 = 0$) in the solution space for $\{U, V, W\}$. It follows, since $\mathcal{U} < 0$, the solution at D2 represents a decaying inhomogeneity.

**Point D3:** Here we obtain the same behaviour as in the case when $k_E \leq k_{EC}$, i.e. D3 is a saddle, corresponding to a decaying density inhomogeneity.

As for region 1 we can now write down explicit solutions of the perturbation equations. In terms of proper time $\tau_E$ we obtain:

$$\Delta = \frac{1}{2c_{\pm E}^2 k_E} + \frac{3}{4} C_0 + \frac{1}{4} C_1 (\frac{1}{t_E} - \frac{1}{t_E})^- + \frac{1}{4} C_2 (\frac{1}{t_E} - \frac{1}{t_E})^+, \quad \mathcal{K} = const = C_k, \quad (7.77)$$

107
where
\[ \xi_t = \frac{1}{3} \sqrt{6 \left( \frac{25}{24} - \frac{k_E^2}{k_{EC}^2} \right)} \]

and \( 0 \leq \xi_t < \frac{1}{3} \).

In this region, none of the solutions correspond to growing modes for density inhomogeneities. The line of stable sinks represent modes where the density perturbation approaches constant value.

**Region 3 -** \( k_E > \frac{5}{2 \sqrt{6} k_{EC}} \). See figures 7.4 and 7.5

**Point D1:** For these values of \( k_E \), the eigenvalues \( \lambda_3 \) and \( \lambda_4 \) are imaginary. Since their real parts are both negative, this solution acts as an attractor, and gives rise to a line of stable spiral points. This repetition of the equilibrium points is, as mentioned before, a consequence of one of the eigenvalues vanishing. Here, the density and magnetic parts of the solution decouple from the curvature (\( W' \) equation), and thus the spiral point exists in any \( W = \) constant plane (see figure 7.5).

**Point D2:** The solutions for \( \mathcal{U} \) are imaginary in this region, so density inhomogeneities oscillates as a sound wave, neither growing nor decaying.

**Point D3:** Here there are no real solutions. Density perturbations again oscillate as sound waves.

This time the solutions are given by
\[
\begin{align*}
\Delta &= C_0 + C_1 \left( \frac{a}{a_E} \right) \left( -\frac{4}{3} + i\beta \right) + C_2 \left( \frac{a}{a_E} \right) \left( -\frac{4}{3} - i\beta \right), \\
B &= \frac{1}{2} \left( \frac{3}{a_E^2} - \frac{3}{2k_E^2} - \frac{3}{6} \right) C_0 \\
&\quad + \frac{4}{3} C_1 \left( \frac{a}{a_E} \right) \left( -\frac{4}{3} + i\beta \right) + \frac{4}{3} C_2 \left( \frac{a}{a_E} \right) \left( -\frac{4}{3} - i\beta \right), \\
\kappa &= C_k.
\end{align*}
\]

with \( \beta = \frac{1}{2} \sqrt{6 \left( \frac{25}{24} - \frac{k_E^2}{k_{EC}^2} \right)} \). In terms of the proper time \( \frac{t}{t_E} \) we obtain:
\[
\begin{align*}
\Delta &= C_0 + \left( \frac{t}{t_E} \right) - \frac{3}{2} \left( C_3 \cos \left( \frac{2}{3} \beta \ln \left( \frac{t}{t_E} \right) \right) + C_4 \sin \left( \frac{2}{3} \beta \ln \left( \frac{t}{t_E} \right) \right) \right), \\
B &= \frac{1}{2} \left( \frac{3}{a_E^2} - \frac{3}{2k_E^2} - \frac{3}{6} \right) C_0 \\
&\quad + \frac{4}{3} \left( \frac{t}{t_E} \right) - \frac{3}{2} \left( C_3 \cos \left( \frac{2}{3} \beta \ln \left( \frac{t}{t_E} \right) \right) + C_4 \sin \left( \frac{2}{3} \beta \ln \left( \frac{t}{t_E} \right) \right) \right), \\
\kappa &= C_k.
\end{align*}
\]
The spiral behaviour round the stable point is evident from the damped oscillatory solution (see figures 7.4 and 7.5).

In this region, the wavelength of solutions falls below a critical wavelength which is related to the Jeans length with a correction due to the magnetic field (see section 7.6.3, below). Thus, these solutions do not result in growing density inhomogeneities, but in general, oscillate as sound waves. The constant density solution due to the magnetic field (corresponding to D1) represents the only stable solution in this region, and can be seen as a spiral point in figure 7.5. Note that this is the only equilibrium point in this figure, as the other 2 solutions represent pure sound waves.

7.6.3 Magnetized Jeans Length

It is clear from the above discussions that the critical scale $k_{EC}$ which appears in our analysis is not quite the scale which determines the onset of oscillatory behaviour. This is determined by the wavenumber

$$k_{osc} = \frac{5}{2\sqrt{6}} k_{EC},$$

whose corresponding wavelength is:

$$\lambda_{osc} = \frac{2\sqrt{6}}{5} \lambda_{EC} = \frac{2\sqrt{6}}{5} \frac{2\pi}{H_E} \frac{4}{3\sqrt{3}} (1 + 3c_a^2)^{1/2} \quad \approx \quad \frac{16\sqrt{2\pi}}{15} \frac{1}{H_E} (1 + \frac{3}{2} c_a^2).$$

109
This scale is closely related to the *Jeans Length* as it determines the wavelength which divides oscillatory from growing or decaying solutions. It takes the form of a general critical wavelength, modified by a linear factor due to the magnetic field. It is clear that since the magnetic field contributes to the overall pressure, it enlarges the size of regions able to resist gravitational collapse, and so increases the critical scale relative to the non-magnetised case.

In the limiting case \( c_{aE}^2 \to 0 \), we obtain:

\[
\lambda_{osc} \to \frac{16 \sqrt{2} \pi}{15} \frac{1}{H_E}
\]  

(7.83)

which is the corresponding scale for a flat dust-radiation model (with no magnetic field) found by Bruni and Piotrowska in [8].

The scale \( \lambda_{osc} \) found above (7.82) differs considerably from that found in [96] in their analysis of a magnetized dust solutions (see equation 68). In a pure dust model, the magnetic field acts as an effective pressure, allowing for the extraction of a "magnetic Jeans length". This length is directly proportional to the Alfvén speed, and vanishes if there is no magnetic field. Furthermore it is considerably smaller (of order \( c_{aE}^2 \)) compared to the critical length found by us using a two-fluid model, since in this case, the magnetic field is not the only source of pressure; there is a contribution due to the radiation component.
Chapter 8

Summary of Part I

In the first two chapters of Part I, we introduced the covariant approach to cosmology and applied it to cosmological perturbations. We showed the derivation of the exact evolution equations for the basic inhomogeneity variables, and illustrated an order-of-magnitude linearization scheme by linearizing these equations about an FRW background.

In chapter 4 and 5 we introduced some basic ideas of dynamical systems theory, and illustrated them by means of two simple perturbation examples. These examples involved setting up and analysing a multi-dimensional dynamical system from a system of linear perturbation equations.

In chapter 6 we motivated the inclusion of magnetic fields as an ingredient in structure formation scenarios. In chapter 7, we synthesized the notation, concepts and analytical techniques presented in the previous chapters in order to analyse magnetized density perturbations as follows:

Assuming a large-scale homogeneous magnetic field, we used the exact general relativistic propagation equations given in [97] to describe the evolution of density and magnetic field inhomogeneities and curvature perturbations. We used the usual approximations of small magnetic field energy density \((B^2/\mu << 1)\), and infinite conductivity \((\sigma \to \infty \Rightarrow E \to 0)\), to simplify these equations. We carefully linearized the exact equations using a two-parameter approximation scheme and found an extra linear term (compared to [97, 98]) in the propagation equation for the spatial curvature \(\mathcal{K}\). This extra term however makes no qualitative difference to the dynamics.

Rather than attempting to solve these equations analytically, we followed Bruni and Piotrkowska [8] (see also chapter 5) by setting up the linear perturbation equations as a four-dimensional autonomous dynamical
system. This approach provides an elegant description of the dynamics of gravitational and magnetic inhomogeneities for a two component medium comprising of pressure-free matter (dust) and radiation, interacting only through gravity.

The equilibrium points for the three-dimensional invariant sets representing the radiation \((w = \frac{1}{3})\) and dust \((w = 0)\) eras correspond to the approximate solutions of the perturbation equations derived in Tsagas and Maartens [96]. For example, in the radiation era we found the following solution for density perturbations on scales much larger than the Hubble radius:

\[
\Delta = C_0 + C_1 \left( \frac{a}{a_E} \right)^{-4\gamma_2^2} + C_2 \left( \frac{a}{a_E} \right)^{(-1 + 2\gamma_2^2)} + C_3 \left( \frac{a}{a_E} \right)^2. \tag{8.1}
\]

The adiabatic decaying mode decays even less rapidly due to the magnetic field, and the decaying non-adiabatic mode decays slightly faster than the corresponding solution in [96]. To zeroth order in \(c_{aE}^2\), these solutions correspond exactly to those given in [96] (see equations (50) and (54) in their paper).

The extra term found in the linearization of the curvature propagation equation is proportional to \(c_{aE}^2\), and thus tends asymptotically to zero. Since the dynamical systems analysis examines the asymptotic behaviour of these models, this extra term makes no difference in the analysis of the dust era. Similarly, the superhorizon dust solutions do not include the non-adiabatic decaying mode induced by the magnetic field, since the Alfvén speed \((c_{aE}^2 = B^2/\mu)\) also tends to zero at late times.

An important feature arising out of the dynamical systems analysis, is that we obtain three distinct evolution regimes for the perturbation modes. These regimes are defined through a critical scale \(\lambda_{EC}\) which is of the order of the Hubble scale at matter-radiation equi-density and includes a modification linear in \(c_{aE}^2\) due to the presence of the background magnetic field:

\[
\lambda_{EC} = \frac{1}{H_0} \frac{8\pi}{3\sqrt{3}} (1 + 3c_{aE}^2)^{1/2}. \tag{8.2}
\]

The three regimes are:

- Large wavelengths - \(\lambda_E > \lambda_{EC}\) These are large-scale density and magnetic inhomogeneities that grow unbounded giving gravitational instability. Solutions for these wavelengths consist of a growing mode, a decaying mode, and a constant mode, represented by a saddle point in the phase space of solutions.
• Intermediate wavelengths - $\lambda_{osc} < \lambda_E < \lambda_{EC}$. These perturbations are over-damped and therefore decay asymptotically to a constant value $C_0$ without oscillating.

• Small wavelengths - $\lambda_E < \lambda_{osc}$. Here the perturbations oscillate like sound waves, while their amplitude decays.

Furthermore, the analysis of the dust models yields a magnetic field corrected scale $\lambda_{osc}$ closely related to the Jeans length, which is a multiple of the general critical scale $\lambda_{EC}$ for flat radiation-dust models:

$$\lambda_{osc} = \frac{2\sqrt{6}}{5} \left[ \frac{1}{H_E} \frac{8\pi}{3\sqrt{3}} (1 + 3c_s^2)^{1/2} \right]$$

$$\approx \frac{16\sqrt{2\pi}}{15} \frac{1}{H_E} (1 + \frac{3}{2}c_s^2), \quad (8.3)$$

where the quantities $c_s^2$, $H_E$, $\lambda_E$ are evaluated at equi-density.

Finally, we note that $\lambda_{osc}$ is more general than the "magnetized Jeans Length" found in [96] for a pure dust model, since it takes into account the pressure effects resulting from a proper two-fluid description.

The dynamical systems approach to perturbation analysis could be extended to non-flat models, which are difficult to analyse quantitatively because the curvature complicates the matter of the gauge invariance of the relevant variables.
Part II
Chapter 9

Uncertainties in Galactic Observational Parameters

9.1 Observational Elements

The essential observational elements are redshift, area distance, and observational volumes [47]. For each source type, the basic observations are apparent size, apparent luminosity (or, equivalently, apparent luminosity and surface brightness) and number counts. A testable relation will relate two of these quantities for a given family of sources. Typically, the galaxy population is quantified via the observational parameters of luminosity, surface brightness, morphology and colour. These are loosely associated with the fundamental properties of mass, angular momentum, evolutionary history and age. Observations of the apparent luminosities of galaxies are often used in order to determine volumes in space which can be used to obtain number counts of galaxies. The luminosity is used as a parameter instead of, and related to, distance, which is not directly observable. In the FRW case, this can be worked out to give expected numbers of various sources in a given volume, which can be used as a test of spatial homogeneity and, in principle, determine cosmological parameters [47].

9.2 Measurement Uncertainties: Luminosity

One of the basic observational parameters for galaxies is apparent luminosity, or apparent magnitude. The intrinsic magnitude of any galaxy is related to its mass, and can only be determined from the measured magnitude if the
distance to the galaxy is known. Other corrections also need to be taken into account, but the major uncertainty comes from the distance measurement.

9.2.1 Distance uncertainties

The lack of depth information in the two-dimensional images of the cosmos which are available to us is a serious limitation on our ability to understand the structures which we perceive. Many of the absolute properties of astronomical phenomena can not be determined without accurate distance information [17]. Accurate distance information is also vital for determining important cosmological parameters such as Hubble constant $H_0$, density $\mu$ of the universe, cosmological constant $\Lambda$ and estimations of the age of the universe.[15].

Redshift distances

The distance of an object down an observers past light cone can be characterized by a distance parameter $\nu$ representing both spatial distance and look-back time. This is directly dependent on the observed redshift $z$. The distance of an object can not be directly obtained from its redshift, as there are local contributions to the redshift due to Doppler shift at the source or the observer. There is no intrinsic way of distinguishing local and cosmological contributions to the redshift $z$. In general, one has [45]

$$ (1 + z) = (1 + z_c) \Pi_i (1 + z_i), \quad (9.1) $$

where $z$ is the observed redshift, $z_c$ is the cosmological redshift and $z_i$ are local contributions to the redshift including Doppler contributions and any 'intrinsic' redshift. Peculiar motions of galaxies fall into this class of local redshift contributions. These arise because the presence of a mass concentration in space will cause a distortion in redshift space along the line of sight. A peculiar velocity, positive or negative, is defined to be the difference between the predicted expansion velocity of a galaxy and its observed velocity. For example, where there is a mean infall of some group of galaxies towards a larger mass concentration (such as the infall of the Local Group to Virgo), then some compression in redshift space relative to physical space will occur as the Hubble expansion velocity is retarded by the gravitationally generated infall. This shows how distortions in redshift space can point us towards locations which have significant peculiar velocities. Further progress, how-
ever, relies on measuring the relative distances of galaxies in order to derive the amplitude of the peculiar velocity [15].

**Extragalactic distance indicators**

For very distant objects, there is no *direct* observational method for measuring or comparing their distances [45]. We thus rely on secondary and higher order distance indicators in order to build the extragalactic distance scale. The use of these indicators requires independent astrophysical information. Most of these indicators are calibrated against the distance to the Large Magellanic Cloud, M31 (the Andromeda galaxy) and M33. The distances to these calibrators are determined using methods based on standard candles, such as the Cepheid period-luminosity relation, and the RR Lyrae absolute magnitude scale. Main sequence fitting of clusters and the luminosity of planetary nebulae are also techniques used for determining the distances to these galaxies.

The five most reliable and well-used extragalactic indicators are [15]:

- **the Tully-Fisher relation**,  
  See section 9.2.2 below.

- **luminosity function of planetary nebulae (PNLF)**  
  There is a sharp cut-off in the maximum luminosity of planetary nebulae within a galaxy. This can be used as a standard candle for distance estimates.

- **surface brightness and luminosity fluctuations**  
  This is a statistical method, originally meant to provide information in the stellar population of a cluster or galaxy at a known distance. The process is inverted and applied to the extragalactic distance scale.

- **supernovae luminosity**  
  As with the PNLF, if the maximum luminosity of supernovae of either type I or II is a constance, they will provide the fundamental determination of the extragalactic distance scale, since they are detectable at very large distances due to their extremely high intrinsic luminosity.

- **globular cluster luminosity function**  
  This rests on the assumption (with very little a priori justification) that the globular cluster luminosity function is universal.
Each of these methods has associated problems and advantages. Most of them have uncertainties associated with the calibration of their zero-points [15]. An detailed examination of the uncertainties relevant in determining the first distance indicator, the Tully-Fisher relationship, follows.

9.2.2 Uncertainties in the Tully-Fisher Relation

This is a relationship which describes the correlation between the width of the HI profile of galaxies (produced by galactic rotation $v_{rot}$), and the absolute magnitude $L$ or diameter $r$ of the galaxies in question. Tully and Fisher [102] mention that the problem previously encountered when trying to establish this relationship was one of observational scatter, due to the difficulty in determining the distances to the calibrating systems, which are necessary in order to obtain accurate absolute magnitudes. Scatter in the Tully-Fisher relation [35] has been claimed to vary from as low as 0.10 mag [20] to as high as 0.7 mag [92, 73].

A good discussion of the sources of scatter in the TF relation is given by Bothun and Mould [16]. They derive total magnitudes by extrapolating from their isophotal aperture-magnitude relation. The error in this measurement is dominated by uncertainty in the calibration zero point, and in the sky background. Line width determinations require knowledge of the galaxy's inclination, and uncertainties in the measured major and minor axes contribute to this. In addition, errors in the measured width are dominated by uncertainty as to where the true reference point lies. When observations are made on clusters of galaxies, all the galaxies are assumed to be at the same distance, and thus the apparent magnitudes can be used to determine the slope of the relation.

Inclination measurements

A galaxy's inclination $i$ is defined as the angle between its equatorial plane and the line of sight. The edge on case thus has $i = 90^\circ$, while a face on galaxy has zero inclination. Inclinations can be determined by examining the velocity field of a galaxy measured in HI. This method yields rather low inclinations compared to those obtained optically [94]. Optically inclination angles $i$ are determined by means of axial ratios, which are usually expressed as the logarithm of major over minor axis, $\log \frac{a}{b} = \log R$. There are different ways to determine this ratio. The most common is to measure the major and minor axes at a certain isophote (see below), and to use Hubble's formula
for an oblate spheroid [94]

\[
\cos^2 i = \frac{q^2 - q_0^2}{1 - q_0^2},
\]

(9.2)

where \( q = \frac{\lambda}{\alpha} \) is the observed minor over major axis, and \( q_0 = \frac{\lambda}{\alpha} \) is the intrinsic axial ratio i.e. thickness over major axis. This is dependent on the type of galaxy observed. Different ways to express this dependence can be found in the literature, but often a mean value of 0.2 is used (see e.g. [82]). A simple law depending on the type of galaxy can also be used to determine \( q_0 \) [18, 54]. Uncertainties in inclination determination are one of the main sources of error in the Tully-Fisher relation. For smaller inclinations, the error in line width due to an uncertainty in inclination is large, due to the fact that line widths are corrected to edge on values.

The galaxies originally used to determine the TF relation all had an inclination exceeding 45° from face-on, resulting in a typical uncertainty in this measurement of 2 – 3% [102].

**Rotational velocity measurements**

There is an uncertainty in their measurements of the HI profile width due to the need to correct for the inclination of the galaxy. Most of the HI in spiral galaxy disks follow circular orbits at the same speed \( v_{rot} \). Since not all of the rotational motion lies along the line-of-sight, different parts of the galaxy have the HI emission Doppler shifted by different amounts. The observed line-of-sight velocity for the HI will lie between \(-v_{rot} \sin i\) and \(+v_{rot} \sin i\) relative to the galaxy’s systemic velocity [17]. This assumes that the galaxy is inclined at an angle \( i \) to the line of sight. Spatially unresolved observations of such disks show the 21cm emission line Doppler-broadened into a characteristic shape showing two sharp peaks at frequencies corresponding to Doppler shifts of \( \pm v_{rot} \sin i \). Thus, measuring the line width gives a direct measure of \( v_{rot} \). Conventionally, this is measured by finding the points in the wings of the line where the intensity has dropped to 20% of the peak value. The difference in Doppler shift between these high - and low - frequency points, \( W_{20} \), provides the measure of the galaxies rotational velocity [17].

A complicating factor in relating \( W_{20} \) to the intrinsic rotation velocity \( v_{rot} \) is that the HI does not follow exactly circular orbits, but also undergoes some slight random motion. This results in a slight broadening of the line profile, so that the \( W_{20} \) measurement tends to overestimate the Doppler
broadening. This can be corrected for by subtracting an estimate of the contribution of these random motions to \( W_{20} \) as follows [17]:

\[
W_R = W_{20} - W_{\text{rand}}.
\]  

For gas with a Gaussian random velocity distribution with dispersion \( \sigma \), the correction to \( W_{20} \) is \( W_{\text{rand}} = 3.6\sigma \) [18]. For faint galaxies, \( W_{20} \) is \( \approx 100 \ \text{km.s}^{-1} \) and \( \sigma \approx 10 \ \text{km.s}^{-1} \), indicating that this correction factor can be highly significant.

Combining the correction for random motion with the correction due to inclination, the rate of rotation can be determined from the observations using:

\[
v_{\text{rot}} = \frac{(W_{20} - W_{\text{rand}})}{\sin i}.
\]  

Sandage [91] shows that there is a clear type dependence in the TF relation when spiral (Sa) galaxies are considered. At a given absolute magnitude, Sa galaxies have the highest rotational velocity of all spiral types.

**Sample biases**

There are basically three kinds of biases in extragalactic samples that can affect distance determinations, namely *sample incompleteness*, the *Scott effect* and the *Malmquist bias* [15]. (Other forms of biasing, such as that engendered by the surface brightness selection effect discussed in section 10.1.3, or effects due to the orientation of galaxies, are not discussed). Of these, the simplest to understand is sample incompleteness. This occurs because sources at a fixed flux limit are only visible to a limited distance. It is easy to correct for the missing faint objects if the rough shape of the distribution of these objects in distance is known.

The *Scott effect* is frequently confused with the Malmquist bias. It comes about when the variable which is being observed, say \( V \), correlates with intrinsic luminosity. At fixed \( V \) there will be a distribution in luminosity. The tendency will be to pick only the brightest objects at fixed \( V \). This introduces a bias into the selection. This is frequently cited as a bias in the Tully-Fisher relation, where fixed \( V \) would correspond to a fixed line width [15].

The *Malmquist bias* occurs when distance errors are involved. Assuming objects are homogeneously distributed in space, more objects at intrinsically large distances will tend to be scattered into the sample than out of it. If the
intrinsic scatter of the distance indicator being used is known, corrections can be made for this bias. For the Tully-Fisher relation, if the intrinsic dispersions is \( \approx 0.4 \text{mag} \) then the correction is 20% in luminosity or 10% in distance [15]. The shortcoming of these corrections is that they only work well when the sample is homogeneously distributed, which is clearly not the case for most galaxy observations.

The \( V_{\text{max}} \) correction (see section 10.1.3) for distance dependent selection effects suffers from this kind of bias [36]. A naive correction would be to take the average \( V_{\text{max}}^{-1} \) weighted with the error distribution of the selection parameters within the selection limits. The probability distribution of the error in the selection parameters must be determined to use this method, which would result in an overcorrection, with objects at the selection limit being counted for only half (assuming a symmetrical error distribution, the other half being outside the selection limits), but an object just outside the selection limits, with a large fraction of its probability function within the limits, would not be included at all. De Jong and Lacy [36] overcome this problem by using a virtual selection limit to away from the original selection limit. They then include objects between the virtual and original selection limits with an appropriate low weight.

Misapplying the corrections for the various biases in extragalactic samples can lead to erroneous answers, particularly when attempting to find distances from samples, e.g. finding the distance to the Virgo cluster by using different astrophysical distance estimates on samples of galaxies in the cluster. It is thus important to fully understand the nature of bias corrections before applying them.

9.2.3 Determining absolute magnitudes

When the distance to a galaxy has been established, the absolute magnitude \( M \) of the galaxy can be calculated. If the distance \( d \) is measured in parsecs, and the apparent magnitude \( m \) is known, then

\[
m - M = 5 \log d - 5.
\]

The quantity \( m - M \) is called the distance modulus of the object.

However, other corrections still have to be applied to the measured absolute magnitudes. A galaxy that contains enough HI for a profile to be measured (to obtain \( v_{\text{rot}} \), as in section 9.2.2 above), also contains large amounts of dust [17]. Some of the luminosity of the galaxy, therefore, will be internally absorbed. Since the dust is concentrated in the galactic disk,
the amount of absorption will be related to the inclination of the galaxy, and will be maximized when the galaxy is seen edge-on. The degree to which the photometry of disk galaxies is affected by dust remains controversial [17]. A number of methods to correct for this effect have been developed which rely on modeling the properties of the dust layer. The impact of the correction can be minimized by obtaining the necessary photometry at infrared wavelengths, since these are not as sensitive to dust absorption effects. This approach was first adopted by Aaronson, Huchra and Mould [1]. On top of this internal absorption effect, any extinction due to the interstellar medium of our galaxy in the direction of observation must be accounted for when determining absolute magnitudes.

If the observed galaxy is moving away from us rapidly, for example galaxies at large distances, the photons emitted in any particular band will have been emitted at shorter wavelengths. Consequently, the observed $V$ magnitude will not be directly connected to the absolute magnitude $M_V$ which we would measure for a stationary object at 10 pc. If the shape of the object's spectrum is known, a relationship between the measured $V$ and $M_V$ can be inferred, where this effect is taken into account by the $K$-correction factor $K$. This is written as [17]

$$m - M = 5 \log d - 5 + A + K.$$  \hspace{1cm} (9.6)

The extra parameter $A$ accounts for interstellar extinction in our galaxy. Both of these effects contribute to dimming the observed magnitude.

9.3 Measurement Uncertainties: Surface Brightness

To a first approximation, the surface brightness of an extended object, like a galaxy, is independent of its distance from us. That is because the solid angle subtended by the object falls off in the same proportion $(1/r^2)$ as the amount of light received from the object. The surface brightness at a given point on the surface of a galaxy is a well-defined, distance-independent quantity. Optical astronomers measure surface brightness in magnitudes per square arcsecond. Surface brightness measurements are complicated by the need to correct for the brightness of the night sky. Small errors in sky brightness can result in large error in the derived surface-brightness profile of a galaxy at large radii [17]. If errors in the combined extrapolation of the sky brightness are comparable to 3% of the measured value of the brightness of the
image, then the uncertainty in the galaxy brightness is of order unity where the surface brightness is of order \(26\mu_B\) (\(\mu_B\) is a \(B\)-band measurement in units of \(\text{mag arcsec}^{-2}\)). Nevertheless, surface brightness data are commonly reported to be reliable even below this value [17].

### 9.3.1 Surface photometry of spiral galaxies

Elliptical galaxies are the simplest systems observed. The photometry of these galaxies is simplified by the fact that they are nearly transparent. The luminosity of ellipticals can therefore be considered to be the unattenuated light from myriads of individual stars [17]. On the other hand, disk galaxies, including spirals, often contain significant quantities of dust, meaning that they are very far from transparent in some wavelengths. This severely complicates the task of deducing their three dimensional structure from the two dimensional images available to us.

### 9.3.2 Dust and inclination effects

In a highly flattened but transparent galaxy, the peak measured surface brightness increases as the galaxy is tipped from face-on to edge-on orientation, since the same luminosity comes from smaller and smaller areas of the sky. In contrast, light from a very dusty galaxy will have been either emitted or last scattered in a thin layer around the galaxy. The surface brightness of a very dusty galaxy will thus be independent of orientation, and its apparent luminosity will be proportional to its projected area, and thus to its inclination [17]. The dependence of photometric parameters, such as scale lengths and total luminosities, on inclination is dependent not only on how much dust a disk contains, but on the distribution of the dust as well.

Since galaxies do not have sharp edges, it is usual to characterize the physical size of a galaxy by quoting an isophotal radius or isophotal diameter. This is the radius or diameter at which some particular surface brightness level is reached, in the absence of dust and inclination. For example, \(R_{25}\) would be the radius one estimates the \(I = 25\mu_B\) isophote to have if the galaxy were seen face on and unobscured by dust.
9.4 Fundamental Plane

9.4.1 Description

For early-type galaxies, a tight empirical correlation is observed to hold, in the local universe, between their central velocity dispersion $\sigma$, effective radius $R_e$, and effective surface brightness $SB_e$. This is known as the Fundamental Plane (FP). It is defined as [34]:

$$\log R_e = \alpha \log \sigma + \beta SB_e + \gamma$$  \hspace{1cm} (9.7)

where Djorgovski and Davis [34] find the parameter values to be:

$$\alpha = 1.39, \quad \beta = 0.3614, \& \gamma = -6.71.$$  \hspace{1cm} (9.8)

The fundamental plane is very thin, indicating that the mechanism for early type galaxy formation is very well defined. It can also be used as a distance indicator, with an accuracy at least as good as the Tully-Fisher relation for spirals. This is useful in the derivation of peculiar velocities. In this capacity, it becomes useful cosmologically as an aid to plotting the large scale velocity field. The fundamental plane is also a useful diagnostic tool for studying the star formation history of elliptical galaxies. Variations of the parameters (9.8) as a function of redshift can be interpreted as general trends in the luminosity evolution of the galaxies [103].

The existence of the FP for early type galaxies leads to the speculation that the two-dimensionality of a set of fundamental properties is a general property of all galaxies [34]. If so, then the Tully-Fisher relation for spiral galaxies may be just one aspect of this hypothetical global relation. Indeed, properties of spirals have been investigated by Bujarrabal et al [19], Whitmore [109] and Watanabe et al [110] who concluded that two principal quantities are necessary to describe the family of spirals. Koda and Sofue [67] claim that a scaling relationship for spirals exists, linking luminosity, radius, and rotation velocity. They investigate possible origins of this relationship. Scaling relationships, such as the Tully-Fisher ($\log I - \log V$, $\log V - \log R$) and Freeman ($\log R - \log L$) relationships, provide observational benefits regarding distance measurements, and theoretical benchmarks to understand the structure, formation and evolution of spiral galaxies.

As well as determining the Tully-Fisher relationship - the fundamental plane would define a Magnitude/Luminosity vs diameter relationship. This relationship plays an important role in determining the whether or not surveys which are complete up to some limiting apparent magnitude $m_*$,
can in fact be said to be complete up to a limiting absolute magnitude $M_\bullet$ [90]. If magnitude and diameter are related strictly to each other, with no scatter, then, in a particular cluster at a particular distance, galaxies would lie along some line in the $(M, \log(r))$ plane (corresponding to projecting the fundamental plane along the $\log(\text{rot})$ axis). This would mean that a cluster catalog which is complete up to some limiting apparent magnitude $m_\bullet$, corresponds to the detection of all galaxies in the cluster with luminosities greater than some intrinsic magnitude $M_\bullet$. If, however, there is a distribution of diameters corresponding to each intrinsic magnitude, then completeness up to some apparent magnitude $m_\bullet$ no longer corresponds (in general) to detecting all galaxies up to any particular intrinsic magnitude $M_\bullet$. Some galaxies which are more intrinsically luminous than the limiting magnitude may not be detectable. This casts doubt on the basis of a luminosity function (e.g. the Schechter function [89]) which does not take these effects into consideration. Some of the drop-off observed in luminosity functions of clusters may be due to surface brightness, rather than magnitude effects.

9.4.2 Determinations of the fundamental plane for spiral galaxies

Moriondo et al [74] determined the parameters of scaling relations analogous to the FP relations for ellipticals for the bulges and discs of 40 spiral galaxies. They used a model consisting of an exponential disk and a bulge and fit these separately to a relation

$$\log R = a \log V + b \log I + c$$

(9.9)

involving a scalelength $R$, a velocity $V$ and a surface brightness $I$. They find that the Tully-Fisher relation for their sample of galaxies is an almost edge-on projection of the disk FP found. Their fit to the disk parameters is very similar to other fits for these parameters for the FP for ellipticals, and they mention that the existence of a “cosmic metaplane” has already been claimed by Bender et al [21]. Moriondo et al find that the accuracy of the disk FP as a distance indicator is comparable but slightly lower than that obtained by the Tully-Fisher relation [74].

Koda et al [67] adopt a different approach, considering global properties of spiral galaxies, namely luminosity, radius and rotation velocity. Their study involves 177 spiral galaxies, and uses the data set presented by Han [58]. They use total $I$-band magnitude $M_I$ (in units of mag), HI velocity
width $W_{20}$ (in units of $km s^{-1}$) and face-on $I$-band isophotal radius $R_{23.5}$ (in units of $kpc$). They find that observed spiral galaxies are distributed on a plane as

$$L \propto (VR)^{1.3} \quad (9.10)$$

and that they are distributed in a surfboard shaped region on the plane.

The two-dimensional scaling relations $L - V, V - R,$ and $R - L$ can be understood as oblique projections of this plane, which Koda et al refer to as the *scaling plane*. The edge-on projection of this plane shows a tighter correlation than the TF projection.

Koda et al mention that the same plane can also be found in the data set of Mathewson, Ford & Buchhorn (MFB) \cite{71}. Figure 9.1 shows a crude plot of the data set of Mathewson and Ford \cite{72}, which includes the earlier MFB data. This shows a trend towards a planar distribution. In generating this picture, all galaxies have been included, and absolute magnitudes are determined from the apparent magnitude and measured velocity as

$$M = m - 5 \log \frac{V}{H} - 25, \quad (9.11)$$

where $V = cmb$ velocity from Table 1 \cite{72} and $H = 75 \ km \ s^{-1}$. 

126
In three dimensions, observed galaxies are spread in the range of the order of 2 for $L$, and several factors for $R$ and $V$. Koda et al [67] hypothesize that the 2-D distribution implies the existence of two dominant physical factors in spiral galaxy formation and that one of the factors is more dominant than the other because of the elongated surfboard shape. They argue that this plane is produced through galaxy formation, and primarily affected by galactic mass and angular momentum. It is important to check whether these limits are real, and relate to the physics of galaxy formation, or whether they are partly the result of detection limits (see section 10.1.3).

Since knowledge of the physics of galaxy formation is essential for the construction of a luminosity function (see section 10.2), the scaling plane could be a useful tool in selecting possible scenarios on which to base the analytic form of the luminosity function. Any hypothesized galaxy formation scenario should be able to reproduce the observed fundamental plane.
Chapter 10

Number Counts and the Galaxy Luminosity Function

10.1 Number Counts

The observed number of any particular type of sources in a sample is a product of their space density and their visibility (detection volume) [80]. Optically selected galaxy samples always have limits in surface brightness in addition to their limits in apparent luminosity and/or angular diameter. The detection volume for a particular type of galaxy in such a survey is then at least a two-parameter function e.g. of luminosity and scale-size [36]. Determining the space density of galaxies from a survey depends on knowing the detection volumes. This means that the only complete description of the galaxy space density which can be obtained observationally is a bivariate distribution function which includes at least two of the three parameters of surface brightness, scale-size and luminosity. This requires knowledge of both distances and surface photometry of a large number of galaxies.

In order to use number counts as a test of space-time geometry, the selection effects on the observations for any systematic catalogue must be understood. These effects depend not only on the surface brightness and apparent luminosity of the sources under consideration, but also on the cosmological distance and intervening matter. A clear understanding of the relationship between the fundamental source properties, and the observed image properties is essential to meaningfully interpret any observations. The properties of the image of any source are determined not only by its intrinsic properties, but also by the imaging and detection system used. It is the
image properties which underlie the detection limits and selection effects of any observation, determining whether a source is 'visible' at all.

The concept of the 'observational map' [45] is useful to examine this problem in a systematic way, based on the equations for the observational map between the "object space" (using coordinates which are source properties such as absolute magnitude, and radius), and the "image space" (with image properties such as apparent magnitude and angular size defining the coordinates). This mapping provides a general framework for the study of selection effects, as well as the effects of various cosmological and astrophysical assumptions on the interpretations of observations.

10.1.1 Defining the observational map

Astronomical observations result in images of astronomical sources (e.g. galaxies) whose nature depends on the choice of instrument. Each image is characterized by a set of image parameters \( I_i \) (e.g. apparent size and magnitude) whose numerical values depend not only on the properties of the source itself, but are also affected by the telescope and imaging system, the distance down the light cone to the source, and the properties of any medium intervening between the source and the detector [45].

The general properties of a class of sources (e.g. spiral galaxies) can be described by source functions \( S_A \), and the characteristics of any particular source in the class (e.g. its luminosity) by source parameters \( s_a \). Similarly, the general properties of the detection system can be described by a set of detector functions \( D_c \), and the characteristics of particular observations (e.g. exposure time) by detector parameters \( d_c \). Finally, the general properties of the space-time and of any intervening medium are described by a set of cosmological functions \( C_B \), with the properties of a particular model specified by cosmological parameters \( c_b \). The results of any specific observation is also dependent on the position of the source relative to the observer, including both the relative direction of the source, and its distance down the observers light cone. The former is easy to specify by means of the polar coordinates \( \theta, \phi \) that determine the null geodesic on which the observation takes pace. The distance down the past light cone can be characterized by a distance parameter \( v \) representing both spatial distance and look back time. The image also depends directly on the observed redshift \( z \). This, unfortunately, consists not only of the cosmological contribution, but also on any peculiar motion or intrinsic reddening of the source. These contributions cannot be isolated (see sec 9.2.1).
The properties of the observed image are dependent on all the above-mentioned functions as follows \[45\]

\[ I_i = F_i(S_A(s_a), D_C(d_c), C_B(v, \theta, \phi, c_b), z). \] (10.1)

The aim of standard observational cosmology is to use the measured \( I_i \) to determine the cosmological parameters \( c_b \). If the source properties and characteristics \( S, s_a \) are unknown, then equation (10.1) cannot be deconvolved. The solution is to seek uniform classes of sources, and to test assumptions on the \( c_b \) by measuring their images in a uniform set of observations. These observations apply the same detector functions \( D_C \) throughout, and assume that each observed source is described by the same set of source functions \( S_A \), with only the parameters \( s_a \) varying for each specific source. Using the appropriate \( C_B^{*} \) and \( c_b^{*} \) for the cosmological model under investigation allows this model to be tested by comparing measured image values to the predicted values, given by \[45\]

\[ I_i = F_i(s_a, d_c, v, \theta, \phi, z) \equiv F_i(S_A^{*}(s_a), D_C^{*}(d_c), C_B^{*}(v, \theta, \phi, c_b^{*}), z). \] (10.2)

In order to understand the direct relationship between the properties of sources and their images, it is appropriate to further limit the function (10.2) by investigating it with fixed detector parameters \( d_c^{*} \), for sources at a fixed position \( P^* = (v^*, \theta^*, \phi^*) \) relative to the observer with fixed redshift \( z^* \). This gives

\[ I_i = F_i^{*}(s_a) \equiv F_i^{*}(s_a, d_c^{*}, v^*, \theta^*, \phi^*, z^*). \] (10.3)

This is what is Ellis, Perry and Sievers \[45\] call the observational map at \( z^*, P^*, \) and \( d_c^{*} \). It is a function from the object representation space (object space) defined by the coordinates \( s_a \) to the image representation space (image space) defined by the coordinates \( I_i \). Where there are two object and two image coordinates, the observational map defines a relation between the object plane and the image plane. For any given source spectrum, the two characteristics which directly describe the radiation distribution are source size, and total energy radiation rate. These can be taken to be the coordinates of the object plane. These correspond to coordinates in the image plane of apparent size and apparent radiation flux for any given observational frequency band (see \[45\] and references therein).

The only information available to the observer is the image characteristics \( I_i \). Problems of interpretation arise when the observational map is
many-to-one; i.e. if the same image characteristics (to within detector sensitivity) can be the result of more than one set of source parameters. Detection limits and selection effects depend on only on the observational variables $I$, particularly the sensitivity limits that determine whether or not it is possible to distinguish between images of two different kinds of objects, or between the same kind of object at different distances. Selection and detection effects are therefore determined by limits in the image space. Due to the non-linearity of the actual maps, significant crowding of the image plane effectively make the map many-to-one, leading to severe interpretation problems, particularly near detection limits in the image space.

By explicitly stating the mapping equations (10.1) in terms of the cosmological parameters, it is possible to directly investigate the observational effects of any cosmological assumptions, and thus avoid the need for recursive procedures that make "corrections" to the observational quantities.

10.1.2 Detection limits

Point Spread Function

One of the main properties of any detection system is the angular response function $S_\nu(\alpha - \alpha')$ of that system (where $\alpha$ is the direction of the observed point relative to the light ray joining the midpoints of the source and detector). In order to form a properly focussed virtual image, the angular response, or point spread function (PSF) must be as nearly as possible frequency independent over the frequency range $[\nu]$ observed. The effective point spread function $S$ is a time average over the exposure period of varying instantaneous images. This atmospheric image variation is known as seeing. As well as causing blurring, these random variations may cause some image degradation. Point spread effects reduce the central surface brightness in images. This is a problem when the brightness of images close to the limits of detection is reduced by point spread to below that limit, thus increasing the number of sources which cannot be detected [45].

The PSF characterizes the effects of seeing by giving the probability density that a photon will hit the imaging device at a point that is offset by a vector $(\alpha - \alpha')$ from where it would have hit in the absence of seeing. The simplest case to consider is a Gaussian PSF. If $\sigma$ denotes the dispersion of the function (or 'radius' of the 'point-spread disk'), then the PSF is given by [45, 17]:

$$S(\alpha - \alpha') = \frac{1}{\pi\sigma^2} \exp\left(-\frac{(\alpha - \alpha')^2}{2\sigma^2}\right).$$  \hspace{1cm} (10.4)
Point spread strongly affects the ability to resolve sources which are close to each other into separate images.

**Absolute detection limits**

The background input $I^B(\nu, \alpha)$ is convolved with the source radiation $I(\nu, \alpha)$ by the detector response $D$. Background input consists of rescattered radiation from other sources (sky brightness) and is assumed to have a random noise spectrum, and not to vary systematically across the source. The point spread background radiation is $I^B(\nu)$. The total measured image intensity is therefore [45]

$$Im^T_{[\nu]}(\alpha) = Im_{[\nu]}(\alpha) + Im^B_{[\nu]}.$$  \hspace{1cm} (10.5)

The ability to distinguish the signal $Im_{[\nu]}(\alpha)$ from the total output $Im^T_{[\nu]}(\alpha)$ depends on the instruments used. After “sky subtraction” we can measure $Im_{[\nu]}$, in the presence of the noise in the total output, down to a detection limit $S_L([\nu])$. This brightness limit determines the apparent size of the object, which is detectable when $Im_{[\nu]} > S_L([\nu])$, but not otherwise. The absolute detection limit occurs when the maximum image intensity drops below $S_L([\nu])$. In practice, an image is only detected when it exceeds this absolute limit by a substantial amount [45]. The night sky, looking towards the galactic pole, and discounting atmospheric and zodiacal contributions, has a brightness of about 23V mag (arc sec)$^{-2}$ [31]. Some detectors also have an upper saturation limit $S_S$ where image brightness can no longer be reliably measured.

**10.1.3 Selection effects**

**Surface brightness selection effects**

Galaxies are detected on the basis of their surface brightness contrast with respect to the night sky background. In 1970, Freeman [53] collated the best photometry available at the time for spiral and irregular galaxies, and discovered that the central surface brightness ($\sigma(0)$) of these systems was virtually constant for almost all disks in the sample (He found in B mag (arc sec)$^{-2}$, $\sigma(0) = 21.65 \pm 0.3$). Disney [31] argued that this apparent physical constant is the result of powerful observational selection effects, due in large part to the limited dynamic range of the photographic emulsions used to record the galaxy samples.
The common value of central surface brightness chosen by catalogued disk galaxies is just such as to give them the largest apparent size when seen against the sky background on a photographic plate [32]. The apparent size of a galaxy of fixed intrinsic luminosity will depend on its surface brightness $\Sigma$. For high $\Sigma$ the galaxy will be intrinsically compact, and for low $\Sigma$, most of the galaxy will be below the detection limits $S_L$. Observed disks have $\Sigma$ close to an optimum value between these two extremes, giving a maximum apparent size. This observed surface brightness is also close to the maximum allowed by internal absorption by dust [32]. The apparent luminosity of galaxy also has a maximum for a particular surface brightness $\Sigma$. We can assume that any sample of galaxies is defined by apparent isophotal luminosity and size limits $l_{iso} \geq l_{lim}$, $r_{iso} \leq r_{lim}$, with a galaxy having to satisfy both criteria simultaneously to be included. Depending on the specific sample limits, the probability of inclusion of such galaxies may be a very sharply peaked function of surface brightness [32].

The recent discovery of low surface brightness (LSB) galaxies, and their inferred large space density is a strong indication that surface brightness selection effects have been severe [15].

**Distance dependent selection effects**

Most field galaxy samples are limited by some readily observable quantity, like apparent magnitude or angular diameter. Since not all galaxies have the same luminosity or physical diameter, they can be seen over different maximum distances before dropping out of a sample due to selection effects. The volume within which a galaxy can be seen and will be included in a sample ($V_{max}$) goes as the distance limit cubed, which results in galaxy samples being dominated by intrinsically bright or intrinsically large galaxies, which have the biggest visibility volume [36]. One way to take account of this effect is to weight each galaxies contribution to the survey by the inverse of its detection volume. This is known as the $V_{max}$ correction method [93]. For a low redshift sample with upper $D_{max}$ and lower $D_{min}$ limits on the major axis angular diameter, the correction is [36]

$$V_{max} = \Omega_f \frac{4\pi}{3} d^{3} \left( \frac{D_{maj}}{D_{max}} \right)^{3} - \left( \frac{D_{maj}}{D_{min}} \right)^{3},$$

with $\Omega_f$ the fraction of the sky used to select the galaxies, $d$ the distance to the galaxies, and $D_{maj}$ the major axis angular diameter of the galaxy.
Redshift or magnitude limits that would limit $V_{\text{max}}$ can be taken into account as well. This correction is unfortunately not unbiased with respect to the spatial distribution of galaxies, as it assumes a uniform distribution. Other methods exist that take density fluctuations into account [46, 111]. In order for the $V_{\text{max}}$ correction to be used, selection parameters must be distance dependent. This leads to what is often called the Malmquist edge bias (see section 9.2.2).

## 10.2 Galaxy Luminosity Function

The distribution of galaxy luminosities is quantified in an analogous manner to that of stars, via the \textit{galaxy luminosity function}, $\Phi(M)$. The quantity $\Phi(M)dM$ is proportional to the number of galaxies that have absolute magnitudes in the range $[M, M + dM]$. This function is conventionally normalized by setting [17]

$$\int_{-\infty}^{\infty} \Phi(M)dM = \nu, \quad (10.7)$$

where $\nu$ is the total number of galaxies per unit volume. The quantity $\Phi(M)dM$ then specifies the number density of galaxies in the magnitude range $(M, M + dM)$.

The determination of the galaxy luminosity function (GLF) is one of the fundamental cosmological observations that can be made [15]. With the GLF, one can estimate the luminosity density of the Universe, as well as the mean mass to light ($M/L$) ratio of the galaxies. The shape of the luminosity function can also help to provide constraints on structure formation theories. The luminosity function provides an extra observation to determine evolutionary and cosmological corrections to number-magnitude relations for galaxies, and can also be used to determine the distances to clusters. Another advantage is that it allows estimation of the frequency of absorption lines in QSO’s due to intervening galaxies.

The usual parameterization which reproduces the observed properties of galaxy surveys is the Schechter function [89]. Defining $\Phi(L)dL$ to be the number density of galaxies with luminosities in the range $[L, L + dL]$, the corresponding function is

$$\Phi(L) = \frac{\Phi^*}{L^*} \left( \frac{L}{L^*} \right) ^{\alpha} \exp \left( -\frac{L}{L^*} \right), \quad (10.8)$$

where $\alpha$ sets the slope of the luminosity function at the faint end, $L^*$ gives the characteristic luminosity above which the number of galaxies falls sharply,
and $\Phi^*$ sets the over-all normalization of galaxy density. This formula was initially motivated by a simple model of galaxy formation. The fit parameters are slightly different for field and cluster galaxies, and with deep surveys, the Schechter function does not correctly predict the number of faint galaxies.

The function is dependent on only one parameter, $L$. However, any galaxy LF is only valid to the surface brightness limit of the survey from which it is derived. As mentioned in section 10.1, a complete description of the galaxy space density needs to be a bivariate distribution function including two of the three parameters of surface brightness, scale-size and luminosity. Bivariate distribution functions are the only proper way to compare samples with different selection criteria, especially when comparing samples at different redshifts. In order to do this, one has to obtain surface photometry and distances of the galaxies under consideration.

Number counts of visible galaxies can suggest a shape for the luminosity function. A simple procedure for estimating $\Phi(M)$ for field galaxies involves measuring the apparent magnitude of all the galaxies in some representative sample. The individual brightnesses are converted to absolute magnitudes by estimating the galaxies' distances. If the redshifts involved are significant, a K-correction to the magnitude may be needed. Finally, the number of galaxies in each absolute magnitude range $(M, M + dM)$ is divided by the volume of space that has been surveyed to convert to a galaxy number density [17].

There are a number of shortcomings to this technique in practice. Various selection effects e.g. Malmquist bias, operate when considering magnitude-limited samples (the usual type of survey) and uncertainties in the distances (see section 9.2.1) affect the derivation of the absolute magnitudes. This analysis also relies on the assumption that the galaxies in any magnitude range $(M, M + dM)$ are uniformly distributed in space. This is an unfounded assumption, as is shown by clustering of galaxies, e.g. the Virgo cluster. For a given absolute magnitude $M$, the number of galaxies found within a volume $V(M)$ can depend as much on the spatial distribution of galaxies in the universe as it does on $\Phi(M)$. This introduces a bias into the estimate for $\Phi(M)$. A variety of techniques for handling this problem are available [22, 46]. The uncertainties in distance and absolute magnitude determinations are discussed in section 9.2.1, and the various selection effects are discussed earlier (see section 10.1.3). A major problem with the procedure described above is that many surveys are described by only one parameter, e.g. magnitude limited. This affects the determination of the 'visibility' or
detection volume of galaxies likely to be included in such a survey.

10.2.1 Visibility of galaxies

As discussed earlier (section 10.1.3), the apparently constant surface brightness of galaxies is the result of observational selection effects, strongly influenced by the brightness of our night sky. The visibility of a galaxy is defined as the volume of space it could be in and still be counted in a sample by exceeding apparent luminosity and angular size limits of the sample [80]. The effect of preferentially selecting galaxies of some 'intermediate' surface brightness works in terms of the apparent surface brightness seen by the observer, and not the intrinsic surface brightness. For nearby, low redshift objects, this distinction is a trivial one. At high redshift, however, K-corrections and cosmological $(1 + z)^4$ dimming will cause the apparent surface brightness of a galaxy to fall dramatically. The galaxies with apparent surfaces brightnesses near the optimum for detection will have higher intrinsic surface brightnesses than those nearby. In any catalogue with fixed angular isophotal diameter, or isophotal magnitudes, galaxies with surface brightnesses near the optimum level will be visible out to greater distances (i.e. will have higher visibility) and so will be preferentially selected. The visibility of a optically-selected galaxy in a survey is therefore at least a two parameter function e.g. of luminosity and scale-size, and depends strongly on these functions.

10.2.2 The two-parameter luminosity function

In view of the above considerations, any galaxy luminosity function, depending as it does on the detection volumes of the galaxies under consideration, should be at least a two parameter function. De Jong and Lacy [36] have investigated the dependence of the local space density of spiral galaxies on luminosity, scale-size and surface brightness. They use a sample of about 1000 Sb-Sdm spiral galaxies, corrected for selection effects in luminosity and surface brightness, as well as for the effects of internal extinction. From this, they derive a bivariate space density distribution in the luminosity-scale-size plane. They assume that the form of the function is that of a distribution in luminosity multiplied by a distribution in scale-size at a particular luminosity. For the luminosity dimension, they assume a Schechter function. Although this assumption is based on observations, the form of the Schechter function itself was originally derived from theoretical considerations, namely,
a self-similar stochastic model for the origin of galaxies [81].

To motivate a form for the distribution in scale-size, De Jong and Lacy consider a disk galaxy formation model, as given by Fall [50]. In this model, disk formation occurs via the collapse of gas within a gravitationally dominant dark matter (DM) halo. The scale-size of a galaxy is determined by its angular momentum, which is acquired, prior to the collapse of the halo, by tidal torques from neighboring objects in the expanding universe. The angular momentum of the system is expressed in terms of a dimensionless spin parameter $\lambda$, which depends on the total angular momentum, the total energy, and the total mass of the system. These quantities are all dominated by the dark matter halo of the galaxy. The spin parameter $\lambda$ is expected to have a log-normal behavior. After some calculation, this can be seen to mean that, at a given luminosity, the simplified Fall and Efstatthiou [51] model predicts the distribution of scale-sizes to be log-normal, and the median value of $r_e$ to vary with luminosity as $r_e \propto L^{-\beta} \approx L^{1/3}$. When combined with the Schechter LF, the full bivariate function for space density as a function of luminosity and effective radius becomes [36]:

$$\frac{d^2n}{dLd\tau_e} = \phi_\star \left( \frac{L}{L_\star} \right)^\alpha \exp \left( \frac{-L}{L_\star} \right) \frac{dL}{L_\star} \times \frac{1}{\sqrt{2\pi\sigma_\lambda}} \exp \left[ -\frac{\ln^2\left( \frac{r_e}{\tau_e} \left( \frac{L}{L_\star} \right)^\beta \right)}{2\sigma_\lambda^2} \right] \frac{dr_e}{r_e}. \quad (10.9)$$

This can be rewritten in terms of absolute magnitudes ($M$) as [36]

$$\phi(M, \log(r_e))dM \, d\log r_e = 0.4 \ln(10) \phi_\star 10^{-0.4(M-M_\star)} \exp(-10^{-0.4(M-M_\star)})dM \times \frac{\ln(10)}{\sqrt{2\pi\sigma_\lambda}} \exp \left[ -\frac{1}{2} \left( \frac{\log(r_e) - 0.4\beta(M-M_\star)}{\sigma_\lambda/\ln(10)} \right)^2 \right] d\log r_e. \quad (10.11)$$

where absolute magnitude $M_\star$ corresponds to luminosity $L_\star$.

The first line in equations (10.9) and (10.11) is the Schechter LF and the second line represents the log-normal scale-size distribution at a given luminosity. In the equations, $\phi_\star$, $\alpha$ and $M_\star$ (or $L_\star$) have the usual meanings for a Schechter luminosity function (10.8), while $r_e$ gives the median disk size for a galaxy with $M = M_\star$, and $\beta$ the slope of the dependence of the median $r_e$ on $L$. The quantity $\sigma_\lambda$ is shown in equations (10.9) or (10.11) to equal the dispersion in $\ln(r_e)$ at a given luminosity.
In order to check the form of their bivariate space density function, De Jong and Lacy [36] use a large sample of spiral galaxies. They correct their sample both for distance dependent biases, and for biases resulting from uncertainties in the selection parameters. In order to correct for the distance dependent selection effects, they use the simple technique of $V_{\text{max}}$ correction. Because the selection parameters are distance dependent, the sample is subject to what is often called the Malmquist edge bias. De Jong and Lacy correct for this by using a virtual selection limit $\sigma$ away from the original selection limit.

Once the two dimensional functional form is checked, and seen to be a good fit to observations, one dimensional projections of the function can be used to determine how these 1D functions depend on limits placed on one of other parameters [36]. Due to selection limits, there are regions in the 2D plane where De Jong and Lacy have no data. A 1D integration would be meaningless in this case. Thus, the real data cannot be used to make the 1D projections. To check how limits in the surface brightness of a distribution can influence the determination of the luminosity function, they integrate the bivariate distribution down to various different surface brightnesses, thus calculating the luminosity function for all galaxies with central surface brightness brighter than the indicated limits. From this they determine that the number of low surface brightness Sb-Sdm galaxies missing from local samples is likely to be small. Correspondingly, the distribution of central surface brightnesses can be integrated down to various limiting absolute magnitudes. The final 1D projection which De Jong and Lacy perform is that of the luminosity density of the local universe as a function of disk central surface brightness.

An accurate bivariate galaxy luminosity function, combined with a good understanding of the observational map (sec 10.1.1) can be used to better understand how selection effects affect our observations. The luminosity function could also be used to constrain galaxy formation scenarios which give rise to it, since the function can be checked against observational data.

### 10.3 Summary of Part II

We have introduced the basic observations of apparent size, luminosity and number counts and described the uncertainties inherent in each type of observation. We particularly examined the uncertainties present in determinations of the Tully-Fisher relation, one of the basic extra-galactic distance
indicators. The fundamental plane for elliptical galaxies was defined, and
determinations of this plane for spiral galaxies were presented. Selection
and detection effects which affect number count observations, particularly
distance-dependent effects were discussed. We introduced the concept of the
observational map in order to clarify the mapping between source and image
parameters. The need for a two parameter galaxy luminosity function was
discussed, and De Jong and Lacey's derivation of a bivariate space density
distribution of galaxies in the luminosity-scale-size plane was presented.

The two parameter galaxy luminosity function, together with the ob-
servational map, can be used to give better insight into the selection effects
operating on galactic observations. Using the ideas of the observational map,
it is possible to investigate whether or not observed limits in the image plane
map (e.g. the limits of the fundamental plane) are due to selection effects,
or are real limits on the source parameters.
Appendix A

Covariant Identities

This identities are written using the notation of [69] and [96], except using \( \nabla \) instead of \( D \) for the orthogonally projected covariant derivative. These identities are used in deriving the propagation equations (assuming a flat background and vanishing cosmological constant):

\[
\begin{align*}
\text{curl} \, \nabla_a f &= -2 \dot{f} \omega_a, \quad (A.1) \\
(a \nabla_a f) &= a \nabla_a \dot{f} + a \dot{f} A_a, \quad (A.2) \\
\nabla^2 \nabla_a f &= \nabla_a (\nabla^2 f) + 2 \dot{f} \text{curl} \omega_a, \quad (A.3) \\
(a \nabla_a J_b) &= a \nabla_a \dot{J}_b, \quad (A.4) \\
\nabla_{[a} \nabla_{b]} V_c &= 0 = \nabla_{[a} \nabla_{b]} S^{cd}, \quad (A.5) \\
\text{div} \, \text{curl} \, V &= 0, \quad (A.6) \\
(\text{div} \, \text{curl} \, S)_a &= \frac{1}{2} \text{curl} \, (\text{div} \, S)_a, \quad (A.7) \\
\text{curl} \, \text{curl} \, V_a &= \nabla_a (\text{div} \, V) - \nabla^2 V_a, \quad (A.8) \\
\text{curl} \, \text{curl} \, S_{ab} &= \frac{3}{2} \nabla_a (\text{div} \, S)_b - \nabla^2 S_{ab}, \quad (A.9)
\end{align*}
\]

where the vectors and tensors vanish in the background and \( S_{ab} = S_{(ab)} \). The magnetic field vector does not vanish in the background, and so its projected derivatives do not commute to linear order. For the magnetic field, the vector identity in (A.5) must therefore be changed to:

\[
\nabla_{[a} \nabla_{b]} B_c = \frac{1}{2} \mathcal{R}_{dcb} B^d - \varepsilon_{abcd} \omega^d B_c, \quad (A.10)
\]

where \( \mathcal{R}_{abcd} \) is the 3 - Curvature tensor formed from \( R_{abcd} \) and the kinematic quantities [97, 98].
Appendix B

Properties of the orthogonal covariant derivative

Assuming the existence of a smooth 4-velocity vector field $u^a$ ($u^a u_a = -1$) at each point of the spacetime, we have a subspace $H_p$ of the tangent space $T_p$ at $p$ which is orthogonal to $u^a$, and $h_{ab} = g_{ab} + u_a u_b$ is the metric in $H_p$. The collection of these subspaces $H_p$ can be called a distribution $D$ or a smooth specification (see [27] and references therein). In general, when the vorticity is non-zero the two vector fields $X^a, Y^a \in D$ satisfy

$$[X,Y]^a = h_{ab} [X,Y]^b = -2u^a \omega_{bc} X^b Y^c,$$

where the defect tensor $D^a_{bc} = u^a \omega_{bc}$ is defined by

$$D^a_{bc} = u^a \omega_{bc}$$

and expresses the fact that the vector $[X,Y]^a$ does not live in $D$. It follows from Frobenius's theorem (see for example Ehlers, 1961 [37]) that $D$ does not possess integrable submanifolds i.e. there are no surfaces orthogonal to the 4-velocity vector $u^a$.

B.1 The spatial derivative

The orthogonal covariant derivative operator $\nabla_a$ is fundamental to the covariant approach to perturbations, since it is through it that all the GI inhomogeneity variables are defined (see section 3.3). It is defined as

$$\nabla_a T^{ab}_{cd} = h^a f h^b g h^c d h^d e \nabla_r T^{r^f}_{pq},$$

141
with total projection on all free indices.

It follows from the above definition that $\nabla_a$ preserves the orthogonal metric $h_{ab}$; that is $\nabla_a h_{bc} = 0$. Indices can be raised and lowered through equations acted on by $\nabla_a$ using $h^{ab}$ and $h_{ab}$. Note that unless the vorticity vanishes, $\omega_{ab} = 0$, $\nabla_a$ cannot simply be used as the covariant derivative of a 3-space, because the defect tensor $D^a_{bc}$ will in general be non-zero. The usual commutation relations must be used with caution. In particular, relation (A.5) must be replaced by (A.10).

### B.2 The 3-curvature tensor

The 3-curvature tensor is defined to be

\[ (3)R_{abcd} = (R_{abcd})_\perp + k_{ad}k_{bc} - k_{ac}k_{bd} \]

\[ \Rightarrow (3)R_{\ell\mu} = \left( R_{\ell\mu} \right)_\perp - 2k^{\mu \nu}_{\ell\mid [\ell k_{\nu]}}, \]

where $\perp$ denotes total projection orthogonal to $u^a$ i.e.

\[ (R_{abcd})_\perp = h^a_\ell h^b_\mu h^c_\nu h^d_\rho R_{\rho\ell\mu\nu}, \]

\[ k_{ab} = \tilde{\nabla}_b u_a = \omega_{ab} + \Theta_{ab} \]  \hspace{1cm} (B.5)

and $\Theta_{ab} = \Theta_{ba} = \sigma_{ab} + \frac{1}{3} \Theta h_{ab}$ is the second fundamental form (or expansion tensor) satisfying $\Theta_{ab} u^b = 0$. When $\omega_{ab} = 0 \Rightarrow k_{ab} = \Theta_{ab}$, $(3)R_{abcd}$ is the 3-curvature of the spaces orthogonal to $u^a$ and will have the usual symmetries of the Riemann tensor in these spaces. In general however when $\omega_{ab} \neq 0$ we have

\[ (3)R_{abcd} = (3)R_{[ab][cd]}, \quad (3)R_{[bcd]} = 2h^{a}_{\ell\mu} \Theta_{\ell\mu\nu} \]

and

\[ (3)R_{\ell\mu} - (3)R_{\mu\ell} = -8\omega^{a}_{\ell\mu} \Theta_{\ell\mu\nu} \]

It follows from the above relations that the corresponding "3-Ricci tensor" is

\[ (3)R_{ac} \equiv (3)R_{abc} = h^{bd} (R_{abcd})_\perp - \Theta k_{ac} + k_{ab}k^b_c, \]

with an anti-symmetric part given by

\[ (3)R_{[cd]} = \frac{1}{3} \Theta \omega_{bc} + (\omega_{db} \sigma^d_c - \omega_{dc} \sigma^d_b), \]

and the "3-Ricci" scalar

\[ (3)R \equiv (3)R^a_a = R + 2R_{bd}u^b u^d - \frac{2}{3} \Theta^2 + 2\sigma^2 - 2\omega^2. \]
On using the Einstein field equations, the above relation gives equation (3.18). Note that the above relations are completely general and can therefore be applied to any spacetime.

From the field equations, the expression for the curvature tensor in a FRW model is ([27], appendix C):

\[
R_{abcd} = \frac{1}{2} (\mu + p) \left( u_a u_c g_{bd} + u_b u_d g_{ac} - u_a u_d g_{bc} - u_b u_c g_{ad} \right) + \frac{1}{3} (\mu + \Lambda) \left( g_{ac} g_{bd} - g_{ad} g_{bc} \right). 
\]  
(B.11)

From this expression it follows that the 3-curvature quantities for a FRW model are:

\[
^{(3)}R_{abcd} = K a^2 \left( h_{ac} h_{bd} - h_{ad} h_{bc} \right), 
\]  
(B.12)

\[
^{(3)}R_{ac} = ^{(3)}R_{ac} = 2 K a h_{ac}, 
\]  
(B.13)

where

\[
K = \frac{a^2}{3} \left( -\frac{1}{3} \Theta^2 + \mu + \Lambda \right), 
\]  
(B.14)

is the usual 3-curvature scalar and the last equality follows from the contracted Bianchi identities for the 3-curvature: \( \bar{\nabla}^a R_{ac} = \frac{1}{2} \bar{\nabla} c R \).
Appendix C

Bianchi and Inhomogeneous Models

C.1 Isometries and Killing vectors

An isometry of a manifold \((M, g)\) is a mapping of the \(M\) into itself which leaves the metric \(g\) invariant [112]. A formal description of this idea involves the Lie derivative as follows. A vector field \(\xi\) generates a one-parameter group of transformations, and vice versa. The orbits of this group are the integral curves of \(\xi\). For these transformations to be isometries, the metric \(g\) should have zero Lie derivative with respect to \(\xi\), i.e.

\[
\mathcal{L}_\xi g_{ab} = 0. \quad (C.1)
\]

This is equivalent to the Killing equation

\[
\xi_{ab} + \xi_{ba} = 0, \quad (C.2)
\]

where \(\xi\) is called a Killing vector field (shortened to KVF).

Cosmological models \((M, g, u)\) can be classified based on the dimension \(s\) of the orbits of the isometry group \(G\), and on the dimension \(d\) of the isotropy subgroup [112]. The dimension \(d\) determines the isotropy properties of the model, while the value \(s\) determines the homogeneity properties.
C.2 Definitions

C.2.1 Bianchi Cosmologies

A Bianchi cosmology \((\mathcal{M}, g, u)\) is a model whose metric admits a three-dimensional group of isometries acting simply transitively on spacelike hypersurfaces, which are surfaces of homogeneity in space-time [112]. The fundamental 4-velocity \(u\) may be orthogonal to the surfaces of homogeneity, or not, leading to, respectively, orthogonal or non-tilted models, and the family of tiled models.

C.2.2 \(G_2\) Cosmologies

A \(G_2\) cosmology is a cosmological model \((\mathcal{M}, g, u)\) which admits an Abelian group \(G_2\) of isometries whose orbits are spacelike 2-surfaces. A \(G_2\) cosmology can thus be used to model spatial inhomogeneities, like density fluctuations or gravitational waves, with one degree of freedom.

C.3 Second order equation for DS analysis of Bianchi I perturbations

A second order DE for \(\Delta_a\) can be derived by eliminating \(Z_a\):

\[
\ddot{\Delta}_a + 2H \dot{\Delta}_a + 2\sigma_a^b \dot{\Delta}_b - \frac{1}{3} \mu \Delta_a - H \sigma_a^b \Delta_b + \sigma_a^b \sigma_b^c \Delta_c - 2S_a = 0. \quad (C.3)
\]

As it stands, the DE (C.3) does not determine the evolution of \(\Delta_a\). The term \(S_a\) must first be dealt with. Various assumptions can be made to do this, the simplest being to consider isocurvature perturbations, which are defined by the condition:

\[
\nabla_a \Delta_a^{(3)} R = 0. \quad (C.4)
\]

This leads to the elimination of \(S_a\), through the Friedmann equation (2.71):

\[
\Delta_a^{(3)} R = 2(\Delta_a^2 + \sigma^2 + \mu). \quad (C.5)
\]

together with the definitions of \(\Delta_a\) (3.12), \(Z_a\) (3.14) and \(S_a\) (5.65) implying that

\[
S_a = 2H Z_a - \mu \Delta_a. \quad (C.6)
\]
Using (5.63), $S_a$ can be expressed in terms of $\Delta_a$ and $\Delta_{(a)}$. This can then be substituted into (C.3) to give [112]:

$$\ddot{\Delta}_{(a)} + 6H \dot{\Delta}_{(a)} + 2\sigma_a^b \Delta_b + \frac{3}{2} \mu \Delta_a + 3H \sigma_a^b \Delta_b + \sigma_a^b \sigma_b^c \Delta_c = 0. \quad (C.7)$$
Bibliography


148


Dynamical systems approach to magnetized cosmological perturbations

Stacey Hobbs* and Peter K. S. Dunsby
Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, Cape Town, South Africa
(Received 7 July 2000; published 17 November 2000)

Assuming a large-scale homogeneous magnetic field, we follow the covariant and gauge-invariant approach used by Tsagas and Barrow to describe the evolution of density and magnetic field inhomogeneities and curvature perturbations in a matter-radiation universe. We use a two parameter approximation scheme to linearize their exact non-linear general-relativistic equations for magneto-hydrodynamic evolution. Using a two-fluid approach we set up the governing equations as a fourth order autonomous dynamical system. Analysis of the equilibrium points for the radiation dominated era lead to solutions similar to the super-horizon modes found analytically by Tsagas and Maartens. We find that a study of the dynamical system in the dust-dominated era leads naturally to a magnetic critical length scale closely related to the Jeans length. Depending on the size of wavelengths relative to this scale, these solutions show three distinct behaviors: large-scale stable growing modes, intermediate decaying modes, and small-scale damped oscillatory solutions.

PACS number(s): 04.30.Nk, 95.30.Sf, 98.80.Cq

I. INTRODUCTION

Magnetic fields play an important role in our Universe. They appear on all scales, from the solar system, through interstellar and extra-galactic scales, to intra-cluster scales of several Mpc. Although magnetic field inhomogeneities have not yet been observed on scales as large as those exhibited by cosmic microwave background anisotropies, it is natural to expect that magnetic fields exist on such scales [1], and that they could play a role in the formation of large-scale structure.

Magnetic fields in intergalactic space are notoriously invisible. In the early 1950s the discovery of synchrotron radiation from the interstellar medium of the Milky Way led to the realization that it possessed a magnetic field. In 1957 Bolton and Wild [2] suggested that the Zeeman splitting of a radio transition should be observable in the interstellar gas. This provided a direct way of measuring the strength of this uniform magnetic field. That almost ten years passed between this suggestion and the first detections of this effect is an indication of the technical difficulty such measurements posed at the time [3]. Since then, advances in observational techniques have produced firmer estimates of magnetic field strengths in interstellar and intergalactic space.

We now know that a large scale, organized magnetic field fills the disk of the Milky Way. Studies of external galaxies indicate that all disc galaxies are permeated by large scale magnetic fields [4], and that \( \mu \)G level fields are common in spiral galaxy disks and halos. Strengths of magnetic fields in intergalactic gas in "normal" galaxy clusters have been measured using Faraday rotation measures (RM) combined with x-ray data. This gives typical field strengths for these fields of between 2 and 6 \( \mu \)G [5], which is comparable to the field strengths in the denser interstellar medium in our galaxy. In 1993, Taylor and Perley [6] discovered that these field strengths were exceeded by those in some less common, but more dense cooling flow clusters, where the field strength can be as high as 30 \( \mu \)G. These fields show ordered components on super-galactic scales [5].

Much of our local universe consists of voids, regions containing very little baryonic matter. As yet, the weaker fields within these voids, which may be relics of true primordial fields, remain unmeasured. Primordial and protogalactic magnetic fields represent the large scale fields which could play a role in structure formation. There are at present over thirty theories about the origin of cosmic magnetic fields at galactic and intergalactic scales. Battaner and Lesch [1] look at astrophysical arguments to examine these models. These can be divided into four main categories, based on when the fields are generated, namely: (a) during inflation, (b) in a phase transition after inflation, (c) during the radiation dominated era, and (d) after recombination. It is the large scale fields which were produced during inflation which are most likely to have implications for structure formation.

Investigating the effect of magnetic fields on structure formation is not a recent endeavor. In 1971, Ruzmaikin and Ruzmaikina [7] gave a Newtonian analysis of the growth of density perturbations in a perfectly conducting medium with magnetic fields. Wasserman [8] considered magnetic influence on galaxy formation, and angular momentum. Kim et al. [9] extend this by including the back-reaction of the fluid on to the field, and as a result, derived a magnetic Jeans length. Battaner et al. [10] present a relativistic analysis of the evolution of magnetic fields and their influence on density inhomogeneities in a radiation dominated universe, while Jedamzik et al. and Subramanian and Barrow [11] have considered magnetic dissipative effects at recombination.

More recently Tsagas and Barrow [12–14] have developed a covariant and gauge-invariant approach to the analysis of magnetized density perturbations, in a universe model containing a perfect fluid. Using the covariant gauge-invariant formalism of Ellis and Bruni [15], they derived a

*Email address: stacey@maths.uct.ac.za
†Email address: peter@vishnu.mth.uct.ac.za
STACEY HOBBS AND PETER K. S. DUNSBY

set of exact non-linear equations for general-relativistic
magneto-hydrodynamic evolution. On linearizing these
equations around a flat Friedmann-Robertson-Walker (FRW)
model, they were able to identify the general relativistic cor-
rections to earlier Newtonian work, including the existence
of a magneto-curvature coupling.

Tsagas and Maartens [16] extended these results, consid-
ering also shape distortion effects due to the fields as well as
density and rotational perturbations. They found explicit so-
lutions to the perturbation equations for the radiation and
dust eras, as well as pure-magnetic density perturbations.
They also identified and analyzed other sources of the mag-
neto effects.

When studying the evolution of density perturbations, the
dynamics of the perturbed universe can be represented by an
autonomous dynamical system, described by a set of coupled
differential equations for the gauge-invariant perturbation
variables [17–19]. This allows the stability behavior of the
model to be investigated with relative ease, with no need to
solve for the background variables. Furthermore, stationary
points of the dynamical system correspond to exact or ap-
proximate analytic solutions of the linearized perturbation
equations, thus providing a useful tool for obtaining solu-
tions in cosmologically interesting situations.

This paper is outlined as follows. In Sec. III, we set up the
evolution equations of the density and magnetic field inho-

geneities, and the curvature perturbations. We discuss in
detail the approximations and linearization procedure in Secs.
IV and V and identify an extra term in the linearized
propagation equation for the spatial curvature \( K \). In Sec. VII,
we set up a five dimensional autonomous dynamical system
equivalent to the linearized propagation equations derived in
Sec. III—these equations give a detailed description of the
full phase space of solutions for perturbation dynamics in a
magnetized dust-radiation universe. The analysis and discus-
sion of this system, with particular emphasis on the three
dimensional invariant sets representing the initial radiation
and final dust dominated states are given in Sec. VII.D. Fi-
nally, conclusions are given in Sec. VIII.

II. DEFINING VARIABLES

A. Spacetime splitting

For all of the following equations, we will use standard
units defined by \( 8 \pi G = c = 1 \). In cosmology, the average ve-
locity of matter at each spacetime event defines a unique
4-velocity vector: \( u^a = dx^a/d \tau \) with \( u_a u_a = -1 \). The funda-
mental fluid-flow lines defined by this vector field are a con-
gruence of worldlines, carrying the fundamental observa-
tors. The projection tensor \( h_{ab} = g_{ab} + u_a u_b \) (with \( h_a h^a = h^a h_a = 3 \)) projects into the local rest spaces of comoving
observers.

The covariant derivative of any tensor \( T \), relative to the
fluid flow, can be split into a time derivative along the fluid
flow, and a totally projected covariant derivative orthogonal to
the fluid flow [20]

\[
\nabla_a T^{ab} = u^a \nabla_a T^{ab},
\]

and

\[
\nabla_a T^{ab} = h^{ab} \nabla_a T^{ab}. \tag{1}
\]
DYNAMICAL SYSTEMS APPROACH TO MAGNETIZED...

\[ E_a = F_{ab} k^b = - F_{ab} u^a, \]  
\[ B_a = \frac{1}{2} \varepsilon_{acd} F^{cd}, \]

where \( E_a u^a = 0 \) and \( B_a d^a = 0 \).

III. EQUATIONS

A. Maxwell’s equations

Using the above decomposition, Maxwell’s equations can be written as follows:

\[ \nabla_b F^{ab} = J^a, \]  
\[ \nabla_{(c} F_{ab)} = 0, \]

where \( J^a \) is the 4-current which generates the electromagnetic field and obeys the conservation law \( \nabla_a J^a = 0 \).

These equations can be split into spatial and temporal parts using the notation defined above:

\[ \text{div} E + 2 \omega^a B_a = q, \]  
\[ \frac{\text{curl} B - \frac{1}{2} \Theta E^a}{\text{curl} B = \frac{1}{3} \Theta B^a}, \]

and

\[ \text{div} B = 2 \omega^a E_a, \]

and

\[ \frac{\text{curl} B - \frac{1}{2} \Theta E^a}{\text{curl} B = \frac{1}{3} \Theta B^a}, \]

B. Gauge-invariant variables

Following [16] and [12], the key covariant and gauge-invariant variables describing inhomogeneity in the fluid and the magnetic field are the comoving spatial gradients of the energy density \( \rho \), the expansion \( \Theta \) and the field density \( B^2 \):

\[ \Delta_a = \frac{\text{d} \widetilde{\nabla} a}{\rho}, \quad \Theta_a = a \widetilde{\nabla} a \Theta, \quad B_a = \widetilde{\nabla} a B^a, \]

where \( B^2 = B_a B^a \). Other basic variables which appear in the exact equations are the spatial gradient of the pressure, \( Y_a = \widetilde{\nabla} a P \), the anisotropic pressure generated by the magnetic field, \( \Pi_{ab} = \frac{1}{2} B^2 h_{ab} - B_a B_b \), and the comoving spatial gradient of the field vector: \( B_{ab} = a \widetilde{\nabla} k B_a \).

C. Medium and propagation formulas

The matter description under consideration is a mixture of dust and radiation which share the same 4-velocity \( u^a \), interacting only through gravity. In general, the evolution equations for density perturbations couple to an entropy evolution equation through the equation of state \( p = p(\rho, s) \), where \( s \) is the entropy density. However, since entropy perturbations are only important on very small scales, we will assume adiabatic perturbations at all times, thus ignoring any entropy contributions [22].

The total fluid equations take the same form as those for a perfect fluid of infinite conductivity. The infinite conductivity approximation allows the electric field to be omitted from Maxwell’s equations while spatial currents are preserved [12].

In this case, Maxwell’s equations (12)–(15) generate three constraints [16],

\[ \omega a B_a = \frac{1}{2} q, \]

\[ \text{curl} B = \frac{1}{3} \Theta B^a, \]

and one propagation equation

\[ B_a = \frac{1}{3} \Theta B^a. \]

D. Conservation equations

Since we are only considering adiabatic perturbations, we only need the total fluid equations. The energy and momentum density conservation equations are respectively [12]

\[ \dot{\rho} + \rho (1 + w) \Theta = 0, \]

and

\[ \rho \left[ 1 + w + \frac{2 \kappa^2}{3 \rho} \right] \dot{Y_a} + \frac{2}{\rho} B_{(ab)} B^{(a)} + \mu^2 \Pi_{ab} = 0, \]

where \( w = p/\rho \). Ignoring entropy perturbations, the pressure can be written as a function of the energy density only, i.e. \( p = p(\rho) \). It follows that the relationship between \( Y_a \) and \( \Delta_a \) is given by

\[ a Y_a = \rho c_s^2 \Delta_a, \]

where \( c_s^2 = dp/d\rho = \dot{\rho}/\dot{\mu} \) is the speed of sound in the fluid.

E. Exact propagation equations for the inhomogeneity variables

The propagation equations for the various spatial gradients are given in [12]. They are, first, the equation for the comoving fractional spatial gradient of the energy density:

124007-3
\[
\Delta_\nu = w \Delta_a - (\sigma_a + \omega_a) \Delta_b - (1 + w) \Theta_a 
- \frac{2 \Theta}{\rho} B_{[ab]} B^b + \frac{2 a \Theta B^2}{3 \rho} u^a + a \Theta u^\nu \Pi_{ba},
\]
(24)

secondly, the equation that governs the spatial gradient of the expansion:

\[
\Theta_\nu(-) = - \frac{\Theta}{3} \Theta_a - (\sigma_a + \omega_a) \Theta_b - \frac{1}{2} \rho \Delta_a 
+ B_b B^b - 3 B_{[ab]} B^b + a \mathcal{R} u_a 
+ \frac{3}{2} a u^b \Pi_{ba} + a A_c - 2 a \nabla_a (\sigma^2 - \omega^2),
\]
(25)

where \( A_a = h_a b A_{[ab]} = \nabla_a A, \quad A = \nabla^a u_a \) and

\[
\mathcal{R} = \frac{1}{4} K + A - 3 (\sigma^2 - \omega^2),
\]
(26)

with

\[
K = 2 (\rho + \frac{1}{3} B^2 - \frac{1}{3} \Theta^2 + \sigma^2 - \omega^2)
\]
(27)

representing the 3-Ricci scalar of the observers instantaneous rest frame.

Finally, the equation describing the evolution of the orthogonal spatial gradient of the magnetic field is given by

\[
a^{-2} h_a ^b h^b (a^2 B_{ab}) = - B_{[ab]} (\sigma_a + \omega_a) + \frac{2}{3} B_a \Theta_b + 2 a B_c \sigma^c [u_b]
- 2 a B_c \omega^c [u_b] + a B^c \nabla_b (\sigma_a + \omega_a) 
- \frac{1}{3} \Theta a (2 B_a u_b + \dot{u}_a B_b)
+ a u^c B_c \left( \sigma_{ab} + \omega_{ab} + \frac{1}{3} \Theta h_{ab} \right)
- a h_a ^b R_{[ab]}^c B^c u^b.
\]
(28)

IV. DETAILED DISCUSSION OF APPROXIMATIONS

In the papers by Tsgas, Maartens and Barrow [14—16], the cosmic magnetic field is treated as a coherent test field propagating on the background. The large-scale magnetic field \( B_a \) is assumed to be too weak to destroy the spatial isotropy of the background spacetime, which is taken to be a flat Friedmann-Robertson-Walker (FRW) model.\(^1\) This is an acceptable physical approximation when the magnetic field density is small compared to the energy density of the fluid i.e. \( B^2 / \rho < 1 \). This issue is discussed in [23], where Zel’dovich calculates that \( B^2 / \rho < 8 \times 10^{-5} \) if the model is to be applicable during the whole of the radiation dominated era. Furthermore, the isotropy of the microwave background can be used to place strong limits on the magnitude of the magnetic field [24—26].

Other authors have used alternative approaches to the problem of maintaining isotropy in the background in the presence of magnetic fields. Battaner et al. [1] show that a mean magnetic field is incompatible with a Robertson-Walker metric and therefore assume that there is no mean magnetic field on cosmological scales, i.e. \( \langle B_a \rangle = 0 \). However, they include the presence of magnetic fields in smaller cells, with field directions random on larger scales. Thus, although there is no mean magnetic field, the model includes an average magnetic energy density \( \langle B^2 \rangle \neq 0 \). Kim et al. [9] also follow this approach, considering that, at recombination, field directions are randomly oriented on scales smaller than the Hubble radius.

In order to maintain the coupling between magnetic irregularities and energy density perturbations in a straightforward way, a perfectly conducting medium is introduced. Looking at the covariant form of Ohm’s law [12], we have

\[
f^a + j_b u^b = \sigma E^a,
\]
(29)

where \( \sigma \) represents the conductivity of the medium. Projecting into the rest space of a fundamental observer yields

\[
f^a = \sigma E^a.
\]
(30)

To have a vanishing electric field, while maintaining non-zero spatial currents \( J^a \neq 0 \), the conductivity of the medium must be infinite, i.e. \( \sigma \to \infty \).

An alternative approach would be to assume a pure magnetic field with no electric field and no spatial currents. This would reduce Eq. (13) to

\[
e^{abc} u_b B_c + \text{curl} B^a = 0,
\]
(31)

which is equivalent to

\[
\dot{u}_{[a} B_{b]} + \frac{1}{a} B_{[ab]} = 0.
\]
(32)

Although this appears to couple the acceleration to the field gradient, when this relation is inserted into Eq. (22), all the magnetic terms cancel out. This effectively decouples the magnetic field from energy density inhomogeneities.

V. APPROXIMATION SCHEME

In order to linearize Eqs. (24), (25) and (28), we introduce two smallness parameters. The first, \( \epsilon_1 \), is used to measure the extent to which the gauge-invariant variables deviate from zero (their value in a flat FRW universe). The other parameter \( \epsilon_2 \) is a measure of the Alfvén speed \( c_s^2 B^2 / \rho \).

\(^1\)Spatial flatness is necessary for the 3-Ricci scalar to be gauge-invariant.
A. Zero-order quantities
The energy density $\rho$, pressure $p$ and expansion $\theta$ do not vanish in the background spacetime. The magnetic field, $B_a$, is treated as a small test field propagating on the background. It follows that these variables can be regarded as zeroth order in our approximation scheme.

B. First-order quantities [in $\epsilon_2$]
In order for the metric in some region of spacetime $\mathcal{U}$ to be written in a perturbed FRW form, the following inequalities must hold for the smallness parameter $\epsilon_2$ [14,27]:

$$\frac{\sigma}{H} < \epsilon_1, \quad \frac{\omega}{H} < \epsilon_1, \quad \frac{|E_{ab}|}{H^2} < \epsilon_1, \quad \frac{|H_{ab}|}{H^2} < \epsilon_1,$$

$$\frac{\nabla_a \rho}{\rho H} < \epsilon_1, \quad \frac{\nabla_a \Theta}{H^2} < \epsilon_1,$$

where $|E_{ab}| = (E_{ab}E^{ab})^{1/2}$, $|\Theta_{ab}| = (\nabla_a D^{ab} \nabla_b)^{1/2}$ etc.

Tsagas [14] extends this definition to a magnetized universe by arguing that closeness to a spatially flat FRW spacetime is maintained when additional restrictions are imposed as follows:

$$\frac{|\Pi_{ab}|}{H^2} < \epsilon_1, \quad \frac{|\nabla_a B_a|}{H B_a} < \epsilon_1 \quad \text{and} \quad \frac{|\Delta_a|}{H^2} < \epsilon_1.$$  \hfill (34)

In an exact flat FRW spacetime, all quantities of order $\epsilon_2$ vanish identically. Thus, $\sigma_{ab}, \omega_{ab}, u_a, \{A, A_a\}, \{E_{ab}, H_{ab}\}, \{\Delta_a, \Theta_a\}, \{B_{ab}, B_a\}$, $\{K_a = a^2 K\}$ are all considered to be first order in $\epsilon_2$.

Note that although the magnetic field vector $B_a$ is considered to be a zeroth order quantity, its magnitude must remain small so that it does not disturb the isotropy of the background. This ensures that the anisotropic pressure generated by the magnetic field, $\Pi_{ab}$, is negligible in the background, and therefore $\Pi_{ab}$ may also be regarded as first order in $\epsilon_2$.

Linearization of the above equations is implemented as follows: All terms higher than first order in $\epsilon_2$ are dropped, as well as terms higher than first order in the Alfvén parameter $\epsilon_2$. Terms like $e_{1,2}$ are however kept.

At the end of the calculation, terms first order in $\epsilon_2$ are dropped relative to zero order terms in the coefficients of quantities that are first order in $\epsilon_1$. This is permissible since the magnetic field is very weak ($B^2/\rho \ll 1$). However, this must be done last, otherwise terms may be dropped relative to others which could later vanish.

VI. LINEAR EQUATIONS
A. Conservation equations
Linearization leaves equation (21) unchanged, but slightly modifies the momentum density equation (22):

$$\rho \left(1 + \frac{2B^2}{3\rho} \right) \dot{u} + Y_a - \frac{2}{a} B_{[ab]} B^b = 0.$$  \hfill (35)

This, together with Eq. (23), gives a useful expression for the acceleration vector:

$$\dot{u} = \frac{1}{1 + \frac{2B^2}{3\rho}} \left[ \frac{2}{a} B_{[ab]} B^b - c_s^2 \Delta_a \right].$$  \hfill (36)

It follows that the divergence of the acceleration $A$ is given by\(^2\)

$$A = \frac{1}{1 + \frac{2B^2}{3\rho}} \left( - \frac{c_s^2}{a^2} \Delta + \frac{B^2}{3\rho} - \frac{B^2}{2pa^2} \right).$$  \hfill (37)

B. Linearized propagation equations
The linearized propagation equations for the key gauge-invariant quantities are obtained by dropping terms of $O(2)$ or higher in both linearization parameters. Note that terms of mixed order ($\epsilon_1 \epsilon_2$) and first order in $\epsilon_2$ are kept to start with (but may be dropped later).

In these equations, the expansion scalar $\Theta$ is replaced by the equivalent expression involving the Hubble parameter i.e. $\Theta = 3H$. Also, to first order, all projected time derivatives of quantities that are first order in $\epsilon_1$ are equal to their normal time derivatives (e.g. $\Delta_{(a)} = \Delta_a$).

It follows that the energy density propagation equation (24) becomes

$$\Delta_a = 3wH \Delta_a - \left(1 + w\right) \Theta_a - \frac{6H}{\rho} B_{[ab]} B^b + \frac{2aHB^2}{\rho} u_a.$$  \hfill (38)

The equation for the expansion gradient (25) involves the spatial gradient of the divergence of the acceleration $A_a$.

This can be determined using expression (37), the identities for commutations between spatial gradients and time derivatives given in Appendix A, and the spatial gradient of the 3-curvature $K$, given by

$$a \nabla_a K = 2p \Delta_a + a B_a - 4H \Theta_a.$$  \hfill (39)

Using these expressions, the evolution of the expansion gradient (to first order in $\epsilon_1$) is

\(^2\)In deriving this, $\dot{\nabla}_a w$ and $\nabla_a c_s^2$ are treated as first order in $\epsilon_1$, as in [13].
\[
\Theta_a = -2H\Theta_a - \frac{1}{2} \rho \Delta_a - \frac{1}{2} a B_a - 3 B_{(ab)} B^b \frac{c_s^2}{1 + w + \frac{2 B^2}{3 \rho}} + \frac{a}{2 \rho (1 + w + \frac{2 B^2}{3 \rho})} \nabla^2 \Delta_a - \frac{2c_s^2}{3 \rho (1 + w + \frac{2 B^2}{3 \rho})} H \Theta_a .
\]

C. Scalar equations

Focusing on the growth, or decay, of density inhomogeneities, the vector field \( \tilde{\Delta}_a \) contains more information than is necessary. We can extract the required information by considering a local decomposition of the spatial gradient of \( \Delta_a \) first introduced in [28]:

\[
\Delta_{ab} = a \nabla_a \Delta_a = W_{ab} + \Sigma_{ab} + \frac{1}{3} \Delta h_{ab} ,
\]

where \( W_{ab} = \Delta_{(ab)} \) represents rotations of the density gradient \( \Delta_a \), \( \Sigma_{ab} = \Delta_{(ab)} - \Delta h_{ab} \) describes the variations of \( \Delta_a \) associated with pancake- or cigar-like structures, and finally \( \Delta = \Delta_a = a \nabla^a \Delta_a \) is related to spherically symmetric gravitational clumping of matter. It is this scalar variable that is important when examining structure formation.

We also need to consider the following complementary scalar variables [13]:

\[
Z = a \nabla^2 \Theta_a , \quad B = \frac{a^2}{B^2} \nabla^2 B^2 , \quad \kappa = a^2 \kappa ,
\]

which represent spatial divergences in the expansion gradient, the energy density gradient of the magnetic field, and perturbations in the spatial curvature respectively.

As before, we linearize the scalar propagation equations by dropping terms of order \( \epsilon_1 \) or \( \epsilon_2 \), but retain all terms that are first order in \( \epsilon_2 \).

Equations describing the propagation of these scalars are

\[
\Delta = 3wH\Delta - (1 + w)Z + \frac{3HB^2}{2\rho} - \frac{HB \kappa}{\rho} - \frac{2c_s^2B^2}{\rho (1 + w + \frac{2 B^2}{3 \rho})} H \Delta ,
\]

\[3\text{This decomposition is analogous to the decomposition of the first covariant derivative of the 4-velocity: Eq. (4).}\]

D. Final linearized system

Combining Eqs. (43)–(46), and finally dropping terms of order \( \epsilon_1 \) with respect to zero order quantities, we obtain a second-order differential equation for the scalar energy density perturbations:
\[
\Delta = -(2 + 3c_t^2 - 6w)H\Delta + \frac{1}{2} (1 - 6c_t^2 + 8w - 3w^2) \rho_\Delta \\
+ c_t^2 \Theta^2 \Delta - \frac{2B^2H}{\rho(1+w)} c_t^2 \Delta - \frac{1}{2} (1 - 3c_t^2 + 2w) \rho c_t^2 B \\
+ \frac{1}{2} c_t^2 \Theta^2 B + \frac{1}{3} (2 - 3c_t^2 + 3w) \rho c_t^2 K. 
\]

It is not immediately obvious that the term containing \(c_t^2\) is negligible with respect to the other terms involving \(\Delta\). In order to clarify this, we can write \(c_t^2\) in terms of \(w\) by noting

\[
c_t^2 = \frac{d(c_t^2)}{dw} = \frac{4(w - c_t^2)}{(1+w)} H. 
\]

In what follows we will consider the above equations for values of \(w \in [0, \frac{1}{3}]\), but more specifically, we will look at the dust \(w = 0\) and radiation dominated \(w = \frac{1}{3}\) cases. From Eq. (58) we see that \(w = 0 \Rightarrow c_t^2 = 0\) and that \(w = \frac{1}{3} \Rightarrow c_t^2 = \frac{1}{3}\). Thus, the term drops away during these era and is negligible at all other times.

Without that term, the second order propagation equation for \(\Delta\) becomes

\[
\Delta = -(2 + 3c_t^2 - 6w)H\Delta + \frac{1}{2} (1 - 6c_t^2 + 8w - 3w^2) \rho_\Delta \\
+ c_t^2 \Theta^2 \Delta - \frac{1}{2} (1 - 3c_t^2 + 2w) \rho c_t^2 B + \frac{1}{2} c_t^2 \Theta^2 B \\
+ \frac{1}{3} (2 - 3c_t^2 + 3w) \rho c_t^2 K. 
\]

Finally, the completely linearized propagation equations for \(B\) and \(K\) are

\[
B = \frac{4}{3(1+w)} \Delta + 4H \frac{(c_t^2 - w)}{(1+w)} \Delta, 
\]

and

\[
K = \frac{4Hc_t^2}{(1+w)} \Delta - \frac{4}{3} \frac{B^2H}{\rho(1+w)} K + \frac{2B^2H}{\rho(1+w)} B. 
\]

Equations (49) and (50) are identical to those in [16], but as a result of slightly different linearization,\(^4\) Eq. (51) has an extra term compared to the corresponding result in [16].

\(^4\)At this point Tsagas and Maartens [16] expand the LHS of Eq. (51) using the definition of \(K\) [see Eq. (42)]; \(K = 2H K + a^2 K\) and then drop the second term on the RHS of Eq. (51) relative to \(2H K\).

VII. DYNAMICAL SYSTEMS ANALYSIS

Instead of attempting to find exact solutions to Eqs. (49)–(51) for a dust-radiation background, we instead follow Bruni and Piotrowska [19] and perform a qualitative analysis of the above system of differential equations.

A. The system

It is useful to change the independent variable from proper time to a function of the scale factor \(a\): with the choice \(\tau = \ln a\) (which yields \(d\tau/dt = H\)), and using the standard harmonic decomposition described in Appendix B, Eqs. (49), (50) and (51) become

\[
\Delta_\tau = -\frac{1}{2} (1 + 6c_t^2 - 15w) \Delta_\tau \\
+ \frac{3}{2} (1 - 6c_t^2 + 8w - 3w^2) \Delta_\tau \\
- c_t^2 \frac{k^2}{a^2 H^2} \Delta_\tau - \frac{3}{2} (1 - 3c_t^2 + 2w) c_t^2 B_{\text{(k)}} \\
- \frac{c_t^2}{2} \frac{k^2}{a^2 H^2} B_{\text{(k)}} + (2 - 3c_t^2 + 3w) c_t^2 K_{\text{(k)}}, 
\]

\[
B_{\text{(k)}} = -\frac{4}{3(1+w)} \Delta_\tau + \frac{4(c_t^2 - w)}{(1+w)} \Delta_\tau, 
\]

and

\[
K_{\text{(k)}} = -\frac{4c_t^2}{3(1+w)} \Delta_\tau + \frac{4c_t^2}{3(1+w)} K_{\text{(k)}} + \frac{2c_t^2}{(1+w)} B_{\text{(k)}}, 
\]

where a prime denotes differentiation with respect to \(\tau\). From now on we will drop the subscript \((k)\).

B. Coefficients in terms of \(w\)

In order to close the system, propagation equations are needed for all background variables. It turns out however that the coefficients of the above differential equations can all be written explicitly as functions of the equation of state parameter \(w\), thus only an evolution equation for \(w\) needs to be found. To achieve this, we normalize the scale factor \(a\) at dust-radiation equi-density by introducing the variable \(S = a a_E\) (see Ehlers and Rindler [29] and Padmanabhan [30]).

Since the two fluids are coupled only through gravity, the energy conservation equation is obeyed separately for each component. This gives \(\rho_E = \rho_g f^2 S^{-3}\) and \(\rho_r = \rho_r f^2 S^{-4}\) for dust and radiation respectively, where \(\rho_E\) is the total energy density at equi-density. Total energy conservation yields \(\rho = (\rho_g f^2)(S^{-3} + S^{-4})\). The only contribution to the total pressure comes from radiation component: \(p = p_r = \rho_r f^2 = (\rho_r f^2) S^{-4}\). It is now straightforward to find \(w\) in terms of \(S\) [19]:
\[ w = \frac{1}{3(S+1)}, \]  
which can be inverted to give
\[ S = \frac{1-3w}{3w}. \]

In this way the expansion of the universe model is parametrized by \( w \), with \( w \in [0, \frac{1}{3}] \) and varies from a pure radiation-dominated (\( t \to 0 \)) to a pure dust-dominated (\( t \to \infty \)) phase.

These results (55) and (56) allow us to write the coefficients that occur in the perturbation equations as functions of \( w \) only (see Appendix C).

The energy density of the magnetic field \( B^2 \) has a radiation-like propagation equation \( \frac{\partial^2 B}{\partial t^2} = (S^n - S^{-4})B^2 \), hence the Alfven speed can be written as
\[ c_a^2 = 6w c_E^2. \]

The propagation equation for \( n \) can now be written down. It is
\[ n' = 3w \left( w - \frac{1}{3} \right). \]

In terms of \( w \), Eq. (52) becomes
\[ \Delta'' = \alpha_\Delta \Delta' + \beta_\Delta \Delta + \gamma_\Delta B + \eta_\Delta K, \]  
where
\[ \alpha_\Delta = -2 \frac{(1-3w^2)}{(1+w)} + \frac{3}{2} \frac{(1+w)}{9}, \]
\[ \beta_\Delta = \frac{3}{2} \left[ 1 + \frac{w^2(5-3w)}{(1+w)} \right] - \frac{8(1-3w)^2}{9(1+w)^2} \]
\[ \gamma_\Delta = -9w c_E^2 \left[ \frac{1-w+2w^2}{(1+w)} \right], \]
\[ \eta_\Delta = 6w c_E^2 \left[ \frac{2 + w + 3w^2}{(1+w)} \right]. \]

Similarly, Eqs. (53) and (54) become
\[ B' = \alpha_B B' + \beta_B \Delta, \]  
and
\[ K' = \alpha_K K' + \beta_K K + \gamma_K B, \]  
where
\[ \alpha_B = \frac{4}{3(1+w)^2}, \]
\[ \beta_B = \frac{4w}{3(1+w)^2}(1-3w), \]
\[ \alpha_K = \frac{16w}{3(1+w)^2}, \]
\[ \beta_K = \frac{8w}{(1+w)c_E^2}, \]
\[ \gamma_K = \frac{12w}{(1+w)c_E^2}. \]

with the evolution of the background determined by Eq. (58).

C. Reducing the order of the system

In the simplest case of a single fluid, Eqs. (59), (61) and (62) together with the condition \( w = \text{const} \), form a fourth-order autonomous system of differential equations. In the more general two-component case, the system is five dimensional due to the inclusion of the propagation equation for \( w \) (58). However, the order of the system can be reduced by introducing a new variable \( U \) [19]. This reduction makes it possible to easily determine the qualitative behavior of \( \Delta \), rather than its exact evolution law. We start by defining \( X = \Delta' \) and
\[ U/X = \Delta' = \Delta''/\Delta, \]
\[ \mathcal{R} = \sqrt{\Delta^2 + X^2}. \]

\( \mathcal{R} \) represents an "amplitude" of the perturbation and it is not directly relevant to our analysis. In order to keep the system dimensionally consistent, we define two new variables,
\[ V = \frac{B}{\Delta}, \quad \text{and} \quad W = \frac{K}{\Delta}. \]

In terms of \( U, V, W \), Eqs. (59), (61) and (62) become
\[ U' = -U^2 + \alpha_\Delta U + \beta_\Delta V + \gamma_\Delta W, \]
\[ V' = \alpha_B U + \beta_B V - V U, \]
\[ W' = \alpha_K + \beta_K W + \gamma_K V - W U, \]

with the coefficients determined by Eqs. (60) and (63), together with the propagation equation for \( w \) (58).

D. Analysis

The system is non-linear, however we can analyze it locally by linearizing about any stationary points, without losing the details of the qualitative dynamics [31,32].

From Eq. (58) we can see that stationary points exist for \( w = \frac{1}{4} \) and \( w = 0 \).

Since Eq. (58) decouples from the rest of the system, the 3-dimensional subsystems corresponding to these values of \( w \) correspond to invariant sets describing the early radiation-
dominated and late dust-dominated periods of dynamical evolution. We now consider these cases separately.

E. Radiation era:  \( w = \frac{1}{3} \)

On substituting \( w = \frac{1}{3} \) into Eqs. (60) and (63), the system for the radiation-dominated era becomes

\[
U' = \frac{U}{3} + U + 2 - 2 c_{a E}^2 V + 4 c_{a E}^2 W,
\]

\[
V' = U - V U,
\]

\[
W' = 1 - 2 c_{a E}^2 W + 3 c_{a E}^2 V - W U. \tag{69}
\]

The equilibrium points for Eq. (69) can easily be determined and are given in Table I.

The 3-dimensional space described by \( \{U, V, W\} \), and \( w = \frac{1}{3} \) (see Fig. 1) is an invariant set, so we can classify the equilibrium points according to the eigenvalues \( \lambda_{1,2,3,4} \) (since \( \lambda_1 = \frac{1}{3} \) is not relevant).

Point R1 is a saddle (not included in Fig. 1 because it is far away from the other points). Orbits close to it may initially evolve towards it, but end up evolving away again. At R1 \( U = 0 \), so density perturbations neither grow or decay. Using Eqs. (64) and (63) it follows that \( \Delta, B \) and \( K \) are constant:

\[
\Delta = \text{const} = C_0.
\]

\[
B = \frac{1}{c_{a E}^2} C_0.
\]

\[
K = -\frac{1}{c_{a E}^2} C_0. \tag{70}
\]

Point R2 is also a saddle, but this time since \( \Delta < 0 \), density inhomogeneities are decreasing. Solutions at R2 again follow from Eqs. (64) and (65):

\[
\Delta = C_1 \left( \frac{a}{a_E} \right)^{-4 c_{a E}^2},
\]

\[
B = \Delta = C_1 \left( \frac{a}{a_E} \right)^{-4 c_{a E}^2},
\]

\[
K = -\frac{1}{2 c_{a E}^2} C_1 \left( \frac{a}{a_E} \right)^{-4 c_{a E}^2}. \tag{71}
\]

Point R3 is a node source, representing unstable equilibrium, so all orbits close to this point evolve away from it. Again since \( \Delta = -1 + 2 c_{a E}^2 < 0 \), solutions at R3 represent a decreasing density inhomogeneity:

\[
\Delta = C_2 \left( \frac{a}{a_E} \right)^{(-1 + 2 c_{a E}^2)},
\]

\[
B = \Delta = C_2 \left( \frac{a}{a_E} \right)^{(-1 + 2 c_{a E}^2)},
\]

\[
K = -\left( 1 + 7 c_{a E}^2 \right) C_2 \left( \frac{a}{a_E} \right)^{(-1 + 2 c_{a E}^2)}. \tag{72}
\]

Point R4 is a stable node, or a sink. Orbits close to this point evolve towards it as \( (a/a_E) \) increases. Since \( \Delta > 0 \) solutions at R4 represent growing density inhomogeneities:

\[
\Delta = C_3 \left( \frac{a}{a_E} \right)^{2},
\]

\[
B = \Delta = C_3 \left( \frac{a}{a_E} \right)^{2},
\]

\[
K = \frac{1}{2} \left( 1 + 2 c_{a E}^2 \right) C_3 \left( \frac{a}{a_E} \right)^{2}. \tag{73}
\]

Solutions at the points given above correspond to approximate solutions (to leading order in \( c_{a E}^2 \)) of the perturbation
The equations during the radiation-dominated era. Therefore by
linearity, the general solutions for the perturbation variables
$\Delta, \beta, \text{ and } \kappa$ are given by a linear combination of Eqs. (70)–
(73):

$$\Delta = C_0 + C_1 \left( \frac{a}{a_E} \right)^{-4c_{aE}^2} + C_2 \left( \frac{a}{a_E} \right)^{-4c_{aE}^2} + C_3 \left( \frac{a}{a_E} \right)^{2c_{aE}^2}$$

$$\beta = \frac{1}{c_{aE}^2} C_0 + C_1 \left( \frac{a}{a_E} \right)^{-4c_{aE}^2} + C_2 \left( \frac{a}{a_E} \right)^{-4c_{aE}^2} + C_3 \left( \frac{a}{a_E} \right)^{2c_{aE}^2}$$

$$\kappa = \frac{1}{c_{aE}^2} C_0 + \frac{1}{c_{aE}^2} \sum \left( \frac{a}{a_E} \right)^{-4c_{aE}^2}$$

This can be written more concisely as

$$\Delta = C_{(0)} + \sum \left( \frac{a}{a_E} \right)^{-4c_{aE}^2}$$

$$\beta = \frac{1}{c_{aE}^2} C_0 + \sum \left( \frac{a}{a_E} \right)^{-4c_{aE}^2}$$

$$\kappa = \frac{1}{c_{aE}^2} \sum \left( \frac{a}{a_E} \right)^{-4c_{aE}^2} + \left( \frac{a}{a_E} \right)^{2c_{aE}^2}$$

(74)

(75)

where $a$ solves the cubic equation:

$$a^3 + (2c_{aE}^2 - 1)a^2 - 2a - 8c_{aE}^2(1 + c_{aE}^2) = 0$$

(76)

and corresponds to the super-horizon solutions given in [16],
with slightly modified exponents, due to the extra term in the
equation for the spatial curvature $\kappa$ (54).

**F. Dust era**

When solving for the equilibrium points in the dust-
dominated era, we substitute $w=0$ into the equations defining
the dynamical system (66)–(68). This yields

$$U' = -U^2 + \frac{1}{2} U + \frac{3}{2} \frac{8}{9} k_E^2 \cdot 2c_{aE}^2 k_E^2 V,$$

$$V' = \frac{4}{3} U - V U,$$

$$W' = -W U.$$

(77)

The equilibrium points of this system are shown in Table II. In
this case, the $U$ and $V$ propagation equations decouple from the
equation for $W'$, which means that the system is effectively only 2-dimensional: A critical scale $\lambda_{EC} = 2 \pi a / k_E$, appears in the analysis of this equation through
its corresponding wave number $k_E^2$:

$$k_E^2 = \frac{27}{16} \frac{1}{3c_{aE}^2} \frac{27}{16} (1 - 3c_{aE}^2).$$

(78)

The solutions at each point will have different behavior depend-
ing on the value of $k_E$. In what follows, we first look at the
general properties of the equilibrium points and their
Corresponding solutions and then give specific information about
their behavior for the three regimes of $k_E$ values that emerge from the analysis.

**Point D1.** These solutions are confined to the $\{U, V\}$ plane
for constant $W$ and approach a constant value independent of
the value of $k_E$:

$$\Delta = \text{const} = C_0,$$

$$\beta = \frac{1}{2c_{aE}^2} \left[ \frac{3}{2} \frac{8}{9} \right] C_0,$$

$$\kappa = \text{const} = C_k.$$

(79)

**Point D2.** This point has different stability behavior for dif-
ferent values of $k_E$. The curvature variable $W$ vanishes inde-
dependently of $k_E$. Since the solution at the point depends on $\xi = \xi(k_E)$, the nature of the solution mode (e.g. growing-
decaying-oscillation) will depend on the value of $k_E$:

$$\Delta = C_1 \left( \frac{a}{a_E} \right)^{-1/4 + \xi},$$

$$\beta = \frac{4}{3} C_1 \left( \frac{a}{a_E} \right)^{-1/4 + \xi},$$

$$\kappa = 0.$$

(80)
DYNAMICAL SYSTEMS APPROACH TO MAGNETIZED ...
sources if $\lambda_3 = 0$ in the solution space for $\{U, V, W\}$. It follows, since $\lambda < 0$, the solution at D2 represents a decaying inhomogeneity.

**Point D3.** Here we obtain the same behavior as in the case when $k_E^2 < k_{EC}^2$, i.e., D3 is a source, corresponding to a decaying density inhomogeneity.

As for region 1 we can now write down explicit solutions of the perturbation equations. In terms of proper time $\tau_{DE}$ we obtain

$$\Delta = C_0 + C_1 \left( \frac{t}{\tau_{DE}} \right)^{(-1/6 + \xi_1)} + C_2 \left( \frac{t}{\tau_{DE}} \right)^{(-1/6 - \xi_1)},$$

$$\beta = \frac{1}{2c_{ed}^2} \left[ \frac{3}{2k_E^2} - \frac{8}{9} \right] C_0 + \frac{4}{3} C_1 \left( \frac{t}{\tau_{DE}} \right)^{(-1/6 + \xi_1)} + \frac{4}{3} C_2 \left( \frac{t}{\tau_{DE}} \right)^{(-1/6 - \xi_1)},$$

$$\kappa = \text{const} = C_k,$$

where

$$\xi_1 = \frac{1}{3} \sqrt{\frac{25}{24} \frac{k_E^2}{k_{EC}^2}}$$

and $0 < \xi_1 < \frac{1}{6}$.

In this region, none of the solutions correspond to growing modes for density inhomogeneities. The line of stable sinks represent modes where the density perturbation approaches constant value.

**Region 3: $k_E^2 > (5/2\sqrt{6})k_{EC}^2$. (See Figs. 4 and 5.)**

**Point D1.** For these values of $k_E^2$, the eigenvalues $\lambda_3$ and $\lambda_4$ are imaginary. Since their real parts are both negative, this solution acts as an attractor, and gives rise to a line of stable spiral points. This repetition of the equilibrium points is, as mentioned before, a consequence of one of the eigenvalues vanishing. Here, the density and magnetic parts of the solution decouple from the curvature ($W$' equation), and thus the spiral point exists in any $W=$ const plane (see Fig. 5).

Point D2. The solutions for $U$ are imaginary in this region, so density inhomogeneities oscillates as a sound wave, neither growing nor decaying.

**Point D3.** Here there are no real solutions. Density perturbations again oscillate as sound waves.

This time the solutions are given by

$$\Delta = C_2 \left( \frac{a}{a_E} \right)^{(-1/4 + i\beta)} + C_3 \left( \frac{a}{a_E} \right)^{(-1/4 - i\beta)},$$

$$\beta = \frac{1}{2c_{ed}^2} \left[ \frac{3}{2k_E^2} - \frac{8}{9} \right] C_2 + \frac{4}{3} C_3 \left( \frac{a}{a_E} \right)^{(-1/4 + i\beta)} + \frac{4}{3} C_2 \left( \frac{a}{a_E} \right)^{(-1/4 - i\beta)},$$

$$\kappa = C_k,$$

with $\beta = \frac{1}{4} \sqrt{6(25 - 2k^2/k_{EC}^2)}$. In terms of the proper time $\tau_{DE}$ we obtain

$$\Delta = C_0 + \left( \frac{t}{\tau_{DE}} \right)^{-1/\kappa} \left[ C_1 \cos \left( \frac{2}{3} \beta \ln \left( \frac{t}{\tau_{DE}} \right) \right) + C_2 \sin \left( \frac{2}{3} \beta \ln \left( \frac{t}{\tau_{DE}} \right) \right) \right],$$

$$\beta = \frac{1}{2c_{ed}^2} \left[ \frac{3}{2k_E^2} - \frac{8}{9} \right] C_0 + \frac{4}{3} \left( \frac{t}{\tau_{DE}} \right)^{-1/\kappa} \left[ C_1 \cos \left( \frac{2}{3} \beta \ln \left( \frac{t}{\tau_{DE}} \right) \right) + C_2 \sin \left( \frac{2}{3} \beta \ln \left( \frac{t}{\tau_{DE}} \right) \right) \right],$$

$$\kappa = C_k.$$

(87)

The spiral behavior round the stable point is evident from the damped oscillatory solution (see Figs. 4 and 5).

In this region, the wavelength of solutions falls below a critical wavelength which is related to the Jeans length with a correction due to the magnetic field (see Sec. VII G, below). Thus, these solutions do not result in growing density inhomogeneities, but in general, oscillate as sound waves. The constant density solution due to the magnetic field (corresponding to D1) represents the only stable solution in this region, and can be seen as a spiral point in Fig. 5. Note that this is the only equilibrium point in this figure, as the other 2 solutions represent pure sound waves.

**G. Magnetized Jeans length**

It is clear from the above discussions that the critical scale $k_{EC}$ which appears in our analysis is not quite the scale which determines the onset of oscillatory behavior. This is determined by the wave number $k_{osc} = (5/2\sqrt{6})k_{EC}$, whose corresponding wavelength is

$$\lambda_{osc} = \frac{2\sqrt{6}}{5} \lambda_{EC} = \frac{2\sqrt{6}}{5} \left( \frac{4}{H_E} \right) \left( 1 + 3c_{ed}^2 \right)^{1/2},$$

$$\sim \frac{16\pi^2}{15} \frac{1}{H_E} \left( 1 + \frac{3}{2} c_{ed}^2 \right).$$

(88)
FIG. 5. Dust dominated era - \( k_{\parallel} > S/2 \sqrt{\kappa_{EC}} \).

This scale is closely related to the Jeans length as it determines the wavelength which divides oscillatory from growing or decaying solutions. It takes the form of a general critical wavelength, modified by a linear factor due to the magnetic field. It is clear that since the magnetic field contributes to the overall pressure, it enlarges the size of regions able to resist gravitational collapse, and so increases the critical scale relative to the non-magnetized case.

In the limiting case \( c_{AE}^2 \rightarrow 0 \), we obtain

\[
\lambda_{esc} = \frac{16\sqrt{2} \pi}{15} \frac{1}{H_E},
\]

which is the corresponding scale for a flat dust-radiation model (with no magnetic field) found by Bruni and Pietrovskaya in [19].

The scale \( \lambda_{esc} \) found above Eq. (88) differs considerably from that found in [16] in their analysis of a magnetized dust solution [see Eq. (68)]. In a pure dust model, the magnetic field acts as an effective pressure, allowing for the extraction of a "magnetic Jeans length." This length is directly proportional to the Alfvén speed, and vanishes if there is no magnetic field. Furthermore it is considerably smaller (of order \( c_{AE}^2 \)) compared to the critical length found by us using a two-fluid model, since in this case, the magnetic field is not the only source of pressure; there is a contribution due to the radiation component.

**VIII. SUMMARY**

Assuming a large-scale homogeneous magnetic field, we use the exact general relativistic propagation equations given in [12] to describe the evolution of density and magnetic field inhomogeneities and curvature perturbations. We use the usual approximations of small magnetic field energy density \( B^2/\rho < 1 \) and infinite conductivity \( \sigma \rightarrow 0 \Rightarrow E \rightarrow 0 \), to simplify these equations. We carefully linearize the exact equations using a two-parameter approximation scheme and find an extra linear term (compared to \( 12,13 \)) in the propagation equation for the spatial curvature \( \kappa \). This extra term however makes no qualitative difference to the dynamics.

Rather than attempting to solve these equations analytically, we follow Bruni and Pietrovskaya [19] by setting up the linear perturbation equations as a four-dimensional autonomous dynamical system. This approach provides an elegant description of the dynamics of gravitational and magnetic inhomogeneities for a two component medium comprising of pressure-free matter (dust) and radiation, interacting only through gravity. The equilibrium points for the three-dimensional invariant sets representing the radiation \( (w = \frac{1}{3}) \) and dust \( (w = 0) \) both correspond to approximate solutions of the perturbation equations derived in Tsagas and Maartens [16]. For example, in the radiation era we find the following solution for density perturbations on scales much larger than the Hubble radius:

\[
\Delta = C_0 + C_1 \left( \frac{a}{a_E} \right)^{-4c_{AE}^2} + C_2 \left( \frac{a}{a_E} \right)^{-1 + 2c_{AE}^2} + C_3 \left( \frac{a}{a_E} \right)^2.
\]

The adiabatic decaying mode decays even less rapidly due to the magnetic field, and the decaying non-adiabatic mode decays slightly faster than the corresponding solution in [16]. To zeroth order in \( c_{AE}^2 \), these solutions correspond exactly to those given in [16] [see Eqs. (50) and (54) in their paper].

The extra term found in the linearization of the curvature propagation equation is proportional to \( c_{AE}^2 \), and thus tends asymptotically to zero. Since the dynamical systems analysis examines the asymptotic behavior of these models, this extra term makes no difference in the analysis of the dust era. Similarly, the superhorizon dust solutions do not include the non-adiabatic decaying mode induced by the magnetic field, since the Alfvén speed \( (c_{AE}^2 = B^2/\rho) \) also tends to zero at late times.

An important feature that arises out of the dynamical systems analysis, is that we obtain three distinct evolution regimes for the perturbation modes. These regimes are defined through a critical scale \( \lambda_{EC} \) which is of the order of the Hubble scale at matter-radiation equidensity and includes a modification linear in \( c_{AE}^2 \) due to the presence of the background magnetic field:

\[
\lambda_{EC} = \frac{1}{H_E} \frac{8 \pi}{3 \sqrt{3} (1 + 3c_{AE}^2)^{1/2}}.
\]

The three regimes are:

- **Large wavelengths:** \( \lambda_{EC} > \lambda_{EC} \). These are large-scale density and magnetic inhomogeneities that grow unbounded giving gravitational instability. Solutions for these wavelengths consist of a growing mode, a decaying mode, and a constant mode, represented by a saddle point in the phase space of solutions.
- **Intermediate wavelengths:** \( \lambda_{esc} < \lambda_{EC} < \lambda_{EC} \). These perturbations are over-damped and therefore decay asymptotically to a constant value \( C_0 \) without oscillating.
- **Small wavelengths:** \( \lambda_E < \lambda_{esc} \). Here the perturbations oscillate like sound waves, while their amplitude decays.
Furthermore, the analysis of the dust models yields a magnetic field corrected scale \( \lambda_{esc} \) closely related to the Jeans length, which is a multiple of the general critical scale \( \lambda_{EC} \) for flat radiation-dust models:

\[
\lambda_{esc} = \frac{2\sqrt{6}}{5} \left[ \frac{1}{H_E} \sqrt{\frac{8\pi}{3}} \frac{1 + \frac{1}{3}c_{sE}^2}{(1 + 3c_{sE}^2)^{1/2}} \right] \\
\approx \frac{16\sqrt{2}}{15} \frac{8\pi}{3} \frac{1}{H_E} \left[ 1 + \frac{3}{2}c_{sE}^2 \right],
\]

where the quantities \( c_{sE}^2 \), \( H_E \), \( \lambda_E \) are evaluated at equilibrium.

Finally we note that \( \lambda_{esc} \) is more general than the "magnetized Jeans length" found in [16] for a pure dust model, since it takes into account the pressure effects resulting from a proper two-fluid description.

**ACKNOWLEDGMENTS**

We thank Mattias Marklund, Christos Tsagas and Roy Maartens for helpful discussions and the NRF (South Africa) financial support.

**APPENDIX A: COVARIANT IDENTITIES**

This identities are written using the notation of [21] and [16], except using \( \nabla \) instead of \( D \) for the orthogonally projected covariant derivative. These identities are used in deriving the propagation equations (assuming a flat background and vanishing cosmological constant):

\[
\text{curl} \nabla_a f = -2f \omega_a, \\
(a \nabla_a f) = a \nabla_a f + af \Lambda_a, \\
\nabla^2 (\nabla_a f) = \nabla_a (\nabla^2 f) + 2f \text{curl} \omega_a, \\
(a \nabla_a J_b \ldots) = a \nabla_a J_b \ldots, \\
\nabla_a (\nabla_b V_c) = 0 = \nabla_a (\nabla_b S_{ab}), \\
\text{div} V = 0, \\
\text{div div} S_a = \frac{1}{2} \text{curl div} S_a, \\
\text{curl curl} V_a = \nabla_d (\text{curl} V)_d - \nabla^2 V_a, \\
\text{curl curl} S_{ab} = \frac{3}{2} \nabla_d (\text{curl} S_{ab})_d - \nabla^2 S_{ab},
\]

where the vectors and tensors vanish in the background and \( S_{ab} = S_{(ab)} \). The magnetic field vector does not vanish in the background, and so its projected derivatives do not commute to linear order. For the magnetic field, the vector identity in Eq. (A5) must therefore be changed to

\[
\nabla_a V_b = \frac{1}{2} R_{dcba} B^d - e_{abcd} \omega^d \tilde{B}_c.
\]

where \( R_{abcd} \) is the 3-curvature tensor formed from \( R_{abcd} \) and the kinematic quantities [12,13].

**APPENDIX B: HARMONIC DECOMPOSITION**

When setting up and analyzing the dynamical system, all the equations can be reduced to ordinary differential equations if we restrict our attention to the harmonic components of the perturbation variables. This is a way of effectively separating the time from the space variables, and involves writing the perturbation scalars in terms of harmonic scalars as follows [14]:

\[
\Delta = \sum_n \Delta_n Q^{(n)}, \\
B = \sum_n B_n Q^{(n)}, \\
K = \sum_n K_n Q^{(n)}.
\]

The scalar harmonics, \( Q_n \), are defined by

\[
Q^{(n)} = 0 \quad \text{and} \quad \nabla^2 Q^{(n)} = -\frac{n^2}{a^2} Q^{(n)},
\]

where \( n = k \gg 0 \) (since we are dealing with a flat model). Spatial flatness of the background also means that \( k \) is simply related to the wavelength \( \lambda \) of the perturbation since \( \lambda = 2\pi a/k \).

**APPENDIX C: COEFFICIENTS IN TERMS OF \( w \)**

\( c_s^2 \) is only formally the speed of sound since the fluids interact only through gravity. Using the energy-density conservation equations for matter and radiation we can express \( c_s^2 \) in terms of the parameter \( w \):

\[
c_s^2 = \frac{\rho}{\rho_E} \cdot \frac{4w}{3(1+w)}. 
\]

The total energy density of the matter-radiation mixture is given by

\[
\rho = \frac{27}{2} \rho_E \frac{w^3}{(1-3w)^4},
\]

where \( \rho_E \) is the energy density at matter-radiation equality. This leads to a similar equation for the Hubble parameter \( H^2 \):
where $H_E$ is the value of the Hubble parameter at matter-radiation equality.

When harmonically decomposing the perturbation equations in Sec. VII, the Laplacian terms give rise to coefficients of the form $k^2/a^2 H^2$. These can be expressed in terms of the variable $w$ as follows:

\begin{equation}
\frac{k^2}{a^2 H^2} = \frac{k^2}{a^2 H_E^2} \cdot \frac{2(1-3w)^2}{3w},
\end{equation}

where $k^2 = k^2/a^2 H_E^2$.