An entropic approach to dynamics

Pedro Henrique Moreira Pessoa

University at Albany, State University of New York, ppessoa@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Physics Commons, and the Radiochemistry Commons

Recommended Citation

https://scholarsarchive.library.albany.edu/legacy-etd/2993

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive.
Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
AN ENTROPIC APPROACH TO DYNAMICS

by

Pedro Henrique Moreira Pessoa

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Arts and Sciences
Department of Physics
Spring 2022
The present thesis deals with different aspects of probability, the method of maximum entropy, information geometry, and dynamical systems and their applications to statistical physics and complex systems. Those topics come together in the framework for dynamics termed entropic dynamics. Work on the topic was initiated by my PhD advisor Professor Ariel Caticha and many colleagues at University at Albany as a method for explaining the dynamical processes in quantum mechanics from first principles of probability theory.

In the first part – chapters 3 and 4 – of the present thesis I will discuss the results of my work on statistical physics of quantum systems, which led to two published articles. In one of these articles I have described the information geometry of Fermi-Dirac and Bose-Einstein statistics. This study addresses an important misconception in the topic: it verifies by direct calculation that the information geometry curvature of a Bose gas does not diverge – rather, it converges to zero – in the phase transition known as Bose-Einstein condensation. This is a counterexample to a long held conjecture that curvature always diverges in phase transitions. In the second article I have discussed how Bose-Einstein condensation, as derived from calculations in the thermodynamic limit, is approached from a finite number of particles. This project involved the development of novel computational tools so that the calculations presented do not rely on thermodynamic limit nor on approximations near critical temperature.

In the second part – chapters 5 and 6 – of the present thesis I will discuss my work on entropic dynamics, presenting the results of two other published articles of mine. The first article establishes the dynamical processes obtained from entropic dynamics. We will observe all the necessary calculations for modeling a system through entropic dynamics and how it is possible to obtain the dynamics of measurable (macroscopic) relevant parameters directly from the insight given by data, regardless of how complex the system's dynamics is at the microscopic level. The second article proposes entropic dynamics as an important method in network sciences. In particular, we discuss how the deviations from “scale free networks” found in real world data can be explain as arising from the entropic dynamical process, making entropic dynamics a promising framework to describe dynamics of networks.
Published work

[1] Pedro Pessoa, Ariel Caticha

Exact renormalization groups as a form of entropic dynamics

*Entropy* 20(1), 25 (2018)


Comment on Tsallis, C. Black hole entropy: a closer look. Entropy 2020, 22, 17

*Entropy* 22(10), 1110 (2020)


Entropic dynamics on Gibbs statistical manifolds

*Entropy* 23(5), 494 (2021)


Entropic dynamics of networks

*Northeast Journal of Complex Systems* 3(1), 4 (2021)

[5] Pedro Pessoa, Carlo Cafaro

Information geometry for Fermi-Dirac and Bose-Einstein quantum statistics

*Physica A* 576, 126061 (2021)

[6] Pedro Pessoa

Entropic dynamics yields reciprocal relations


[7] Pedro Pessoa

Bose-Einstein statistics for a finite number of particles

*Physical Review A* 104, 043318 (2021)
Pedro Pessoa

Legendre transformation and information geometry for the maximum entropy theory of ecology

Bruno Arderucio Costa, Pedro Pessoa

Statistical mechanics of unconfined systems: challenges and lessons
I would like to thank my PhD advisor Ariel Caticha whose remarkable teaching and research enthusiasm served as great inspiration during my years at UAlbany, as our several discussions about the nature of entropy shaped my understanding of science. Moreover, I am grateful to have collaborated with Prof. Caticha into pursuing entropic dynamics.

I would also like to thank the members of my committee: Carlo Cafaro, with whom I had the opportunity to collaborate in the work on information geometry of quantum gases; Herbert Fotso, to whom I worked as an teaching assistant in his quantum mechanics course; Oleg Lunin, who has guided me since my first semester in graduate school when I took his quantum mechanics course; and Daniel Robbins, whose help was fundamental in my scientific endeavours on Bose-Einstein statistics.

I would also like to thank many of my mentors and colleagues whose discussions and guidance helped through my years at both University at Albany and University of São Paulo, especially to Nestor Caticha, Felipe Alves, Mo Abedi, Marcelo Boareto, Nicholas Carrara, Gabriel Cozzella, Felippe Cronemberger, Diego González, Otto Heringer, Selman Ipek, Steve Pressé, and Kevin Vanslette. Moreover, I would like to thank the excellent scientific collaboration efforts from my dearest friends Bruno Arderucio Costa and Felipe Xavier Costa.

Finally, I would like to thank my beloved partner Marsha Rikard. I would not be able to finish this tome without her unconditional support during my many rounds of research, writing, and revision.

I was partially funded by CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico— (scholarship GDE 249934/2013-2).
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>Published work</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENT</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Background</td>
<td>5</td>
</tr>
<tr>
<td>2.1 Probabilities</td>
<td>5</td>
</tr>
<tr>
<td>2.1.1 Boolean Algebra notation</td>
<td>6</td>
</tr>
<tr>
<td>2.1.2 Designing a rational theory</td>
<td>7</td>
</tr>
<tr>
<td>2.1.3 Why probabilities?</td>
<td>8</td>
</tr>
<tr>
<td>2.1.3.1 Disjunction and sum rule</td>
<td>9</td>
</tr>
<tr>
<td>2.1.3.2 Conjunction and product rule</td>
<td>10</td>
</tr>
<tr>
<td>2.1.3.3 Obtaining probabilities</td>
<td>11</td>
</tr>
<tr>
<td>2.1.4 Other discussions in probability theory</td>
<td>12</td>
</tr>
<tr>
<td>2.1.4.1 Independence and mutual exclusivity</td>
<td>12</td>
</tr>
<tr>
<td>2.1.4.2 Marginalization</td>
<td>13</td>
</tr>
<tr>
<td>2.1.4.3 Random variables</td>
<td>14</td>
</tr>
<tr>
<td>2.1.4.4 Expected values</td>
<td>15</td>
</tr>
<tr>
<td>2.2 Entropy</td>
<td>17</td>
</tr>
<tr>
<td>2.2.1 Caticha’s design criteria</td>
<td>18</td>
</tr>
<tr>
<td>2.2.2 Gibbs distribution/Exponential Family</td>
<td>21</td>
</tr>
<tr>
<td>2.2.3 MaxEnt and thermodynamics</td>
<td>23</td>
</tr>
<tr>
<td>2.2.3.1 Second law of thermodynamics</td>
<td>23</td>
</tr>
</tbody>
</table>
5. Entropic dynamics on Gibbs statistical manifolds
   5.1 Introduction
   5.2 Entropic Dynamics
       5.2.1 Change happens
       5.2.2 Transition probabilities
       5.2.3 The prior
       5.2.4 The constraint
       5.2.5 Maximizing the entropy
   5.3 Transition on Curved Spaces
       5.3.1 Normal coordinates
       5.3.2 Changing to original coordinates
   5.4 Entropic Time
       5.4.1 Introducing time
       5.4.2 The entropic arrow of time
       5.4.3 Calibrating the clock
   5.5 Diffusion and the Fokker-Planck equation
       5.5.1 Derivatives and divergence
   5.6 Specific Consequences: entropic dynamics and reciprocal relations
   5.7 Examples
       5.7.1 General Gaussian Manifold
       5.7.2 Gaussian submanifold around an entropy maximum
       5.7.3 2-Simplex Manifold
   5.8 Conclusions
6. Entropic dynamics of networks
   6.1 Introduction
   6.2 Gibbs distributions of graphs
## LIST OF FIGURES

2.1 A graphical representation of the conditions for extensivity. On a system composed of several subsystems each sufficient statistic for the composite system (in the gas example energy, volume and number of particles for each molecule species) will be approximately the sum of a analogous sufficient statistics calculated at each subsystem. In physics this will be violated, for example, when the boundary energy or long-range interactions are not negligible even in the thermodynamic limit. .......................................................... 29

3.1 Dimensionless quantity $\tilde{g}_f$, related to metric determinant as (3.18) for fermions. Figure previously published in [5]. .......................................................... 49

3.2 Dimensionless quantity $R_f$, related to metric determinant as (3.21) for fermions. Figure previously published in [5]. .......................................................... 49

3.3 Countour plot of $R$ in (3.21) as a function of $\xi$ and $\beta$ – in units of $\kappa = 1$ – for fermions in the values of $\eta = 1/2$ and $\eta = 2$. Figure previously published in [5]. 51

3.4 Dimensionless quantity $\tilde{g}_b$, related to metric determinant as (3.28) for different values of $\beta$ (in units of $\kappa = 1$) for bosons in the values of $\eta = 1/2$ and $\eta = 2$. Figure previously published in [5]. .......................................................... 53

3.5 Dimensionless quantity $R_b$, related to curvature as (3.30) for different values of $\beta$ (in units of $\kappa = 1$) for bosons in the values of $\eta = 1/2$ and $\eta = 2$. Figure previously published in [5]. .......................................................... 53

3.6 Countour plot of curvature, $R$ in (3.30), compared to their equivalents without the ground state terms, $R_0$ in (3.24). Represented as a function of $\xi$ and $\beta$ (units of $\kappa = 1$) in the relevant region for bosons and for the values of $\eta = 1/2$ and $\eta = 2$. Figure previously published in [5]. .......................................................... 54

4.1 Fraction of particles in the ground state (4.25) for $\eta = 1/2$ and $\eta = 2$. On both cases it can be seen how the fraction of particle in the ground state, for finite $N$, approach, smoothly, the non-analytical curve for the thermodynamic limit. Figure previously published in [7]. .......................................................... 71
### 4.2 Specific heat of a Bose gas (4.26) for $\eta = \frac{1}{2}$ and $\eta = 2$. In the first picture it is seen that the quantity is continuous, albeit not smooth, as $\beta$ approaches $\beta_c$ in the thermodynamic limit. In the second one it is seen how in the thermodynamic limit, the value diverges is discontinuous when $\beta$ approaches $\beta_c$ from the left. On both cases it can be seen how the non-analytical curve is approached from the curves of specific heat for finite $N$. Figure previously published in [7].

### 4.3 Graphs for $(\frac{\partial c_v}{\partial T})_{N,\kappa}$ of a Bose gas (4.28) for $\eta = \frac{1}{2}$ and $\eta = 2$. On the first image it is seen that the quantity becomes smaller than the one expected from calculations in the thermodynamic limit. On the second one it is seen how in the thermodynamic limit, the value diverges to $-\infty$ when $\beta$ approaches $\beta_c$ from the left. Figure previously published in [7].

### 4.4 Graph for the value of $z$, defined in (4.30), calculated numerically for $N$ ranging from $10^2$ to $10^8$ (scattered blue points) and the approximation in order of $N^{-1/3}$ (solid orange line) — meaning $z(N) = z_m + \bar{z} N^{-1/3}$ as in (4.30). Figure previously published in [7].

### 4.5 Graphs for $c_v$ in (4.26) and $(\frac{\partial c_v}{\partial T})_{N,\kappa}$ in (4.28) of a Bose gas for $\eta = \frac{1}{4}$. Unlike in the previous pictures, we see that both quantities are smooth for this value of $\eta$. Figure previously published in [7].

### 5.1 Drift velocity field for the entropy maximum example. Figure previously published in [3].

### 5.2 Unbounded stationary probability for the entropy maximum example. Figure previously published in [3].

### 5.3 Drift velocity field for the 2-simplex in (5.58). Figure previously published in [3].

### 5.4 Static invariant stationary probability for the three-state system. Figure previously published in [3].

### 6.1 Degree distribution derived from the solution of the IVP (6.27) for different values of $L = 2^9$ (left), $2^{13}$ (center), and $2^{30}$ (right). The degree distribution does not depend on $dp_0$ and have a very similar behaviour across $L$. Figure previously published in [4].

### 6.2 Re-scaled degree distribution for the IVP (6.27) irrespective of the number of links $L$. The result fits well with a Weibull and Gamma, also known as power law with cutoff, distributions within most of the allowed range. Figure previously published in [4].
6.3 Degree distribution derived from the solution of the IVP (6.28). The value of $dp_0$ does not interfere with the degree distribution. The result fits well with a Weibull distribution for a network with many connections, $L > 10^4$. Figure previously published in [4].

B.1 Graph for the values of $\gamma^*$ (above) and $\lambda_2^*$ (below) calculated numerically for $N$ ranging from $10^2$ to $10^5$ (scattered blue points) and the approximation in order of $N^{-1/3}$ and $N^{-2/3}$ (solid orange line) respectively — meaning $\gamma^*(N) = \bar{\gamma}N^{-1/3}$ and $\lambda_2^*(N) = \lambda_2N^{-2/3}$ as in (B.3), with $\lambda_2$ and $\bar{\gamma}$ given by (B.24). Figure previously published in [7].
CHAPTER 1

Introduction

In the preface to the Russian edition of Landau and Lifshitz’s book in Statistical physics [10] it is said:

“It is a fairly widespread delusion among physicists that statistical physics is the least well-founded branch of theoretical physics. (...) , Gibbs provided a general method, which is applicable in principle to all problems that can be posed in statistical physics, but which unfortunately has not been adequately taken up. The fundamental inadequacy of the majority of existing books on statistical physics is precisely that their authors, instead of taking this general method as a basis, give it only incidentally. Statistical physics and thermodynamics together form a unit. All the concepts and quantities of thermodynamics follow most naturally, simply and rigorously from the concepts of statistical physics.”

The cited method developed by Gibbs [11] consists on assigning probabilities distributions to the (micro)states of a physical system by maximizing the entropy under expected value constraints for the conserved quantities of their Hamiltonian (classical or quantum) dynamics. The distributions obtained this way are known as Gibbs distributions or Gibbs measures.

The quoted text is credited as written in 1937[1]. Therefore, it predates (i) the revolutionary work of Claude Shannon [12] - which establishes entropy as a fundamental concept in information theory - (ii) the work of Jeffreys [13] and Cox [14] - stating probability as a tool for logical reasoning with incomplete information, (iii) the work of E.T. Jaynes [15, 16] - that uses the maximization of Entropy as a sole foundation of thermodynamics -. and, (iv) the work of Giffin and Caticha [17] that explains how Bayesian inference methods are an application of maximum entropy. Time has shown that the original insight of the Russian authors was not only correct, but an understatement. As of today, the ideas of Gibbs have evolved into the method of maximum entropy (MaxEnt), which is not only the main

[1] The book from which the quoted text is an English version published in 1980. But this preface is dated from 1937-39 as part of the original Russian editions.
foundation of thermodynamics, but a foundation for the whole field of entropic inference [18, 19, 20, 21, 22]. It is not a surprise, therefore, that several authors have applied maximum entropy to a plethora of subjects far from thermodynamics. To cite a few: economics [23, 24], ecology [25, 26], cell biology [27], opinion dynamics [28, 29] and several others topics [30, 31, 32, 33]; particularly, MaxEnt also provides a foundation for network sciences [34, 35, 36, 37].

It does not predate, however the work of Fisher developing the primordium of what is now known as classical statistical theory on the concept of sufficiency. Namely a statistic (function of the sampled values) is sufficient when, in the words of Fisher [38], “no other statistic which can be calculated from the same sample provides any additional information as to the value, of the parameter to be estimated.”. It was later proven, independently by Darmois [39], Pitman [40] and Koopman [41] that the only distributions for which a finite number of sufficient statistics exists independently of the sample size are the distributions in the exponential family, which coincides with Gibbs distributions. Hence, if a probability can be estimated in terms of classical statistics it is also a result of some form of MaxEnt.

Even though Landau and Lifshitz are some of the most praised authors in theoretical physics, almost a century later – and despite all posterior work in probability theory and theoretical physics confirming it – their original words on statistical physics remain vastly underappreciated. Widely used textbooks in statistical physics written after [10], e.g. [42, 43, 44, 45], insist on giving the maximum entropy method “only incidentally” instead of “taking this general method as a basis”. More than this, they seem to work on the impression that statistical physics needs to be constructed to match thermodynamics, while to adequately take up the work of Gibbs one should realise that “All the concepts and quantities of thermodynamics follow most naturally, simply and rigorously from the concepts of statistical physics”.

The present thesis will focus on two major topics: (i) the consequences of the MaxEnt formalism when applied to quantum gases, and (ii) the extension of MaxEnt into a direct formalism for dynamical systems. The following chapter gives a brief review of probability theory and MaxEnt, as well as a brief introduction to the topic of information geometry (IG). This background discussion will focus on the results that will be found relevant in the subsequent chapters, while some focus will be given to the interpretation of these topics.
Chapter 3 presents the result of an investigation — made in collaboration with C. Cafaro [5]. This work recognizes the shortcomings of previous works on the subject [46, 47, 48]. In particular, it was presented in those works that the geometric curvature of a system of Bosons diverges in the phase-transition known as Bose-Einstein (BE) condensation. Our work notices that, when the particles in the ground state are appropriately taken into account, the divergent behaviour disappears and curvature, instead, converges to zero. This result is an explicit counterexample to the long held conjecture that curvature always diverges at a phase transition [49, 50, 51, 52].

Chapter 4 presents a similar discussion on BE statistics [7]. The grand-canonical description of a gas of bosons in the thermodynamic limit is presented in several textbooks on statistical physics. However, previous studies of BE statistics for a finite number of particles have relied on approximations done in the vicinity of the thermodynamic limit [53, 54, 55, 56, 57, 58, 59, 60]. In the work presented in [7] and chapter 4, the calculation of the thermodynamic quantities that indicate phase transitions — namely, the fraction of particles in the ground state and the specific heat — are calculated for an arbitrary number of particles. These calculations require inverting the polylogarithm family of functions. To the best of my knowledge, there is no analytic expression for inverse polylogarithms however, methods for calculating the values of inverse polylogarithms numerically are available in my GitHub repository [61]. Graphs for the calculated thermodynamic quantities are presented in comparison to the results previously obtained in the thermodynamic limit. In particular, it is observed that for the gas trapped in a 3-dimensional box the derivative of specific heat reaches smaller values than what was expected in the thermodynamic limit. This result is also verified with analytical calculations.

Chapter 5 presents the work done in collaboration with F. X. Costa and A. Caticha published in [3]. This investigation extends the entropic dynamics (EntDyn) formalism — previously used to derive quantum mechanics [62, 63] and quantum field theory [64] — into a general framework for dynamics beyond physics. As MaxEnt involves updating probability distributions when new information arrives, EntDyn obtains the transition probabilities — or alternately the laws of dynamics — with the known information about how the system changes. In EntDyn time is an emerging parameter — the system is its own clock. Under the general assumption of continuous motion in a manifold of Gibbs distributions, the resulting...
dynamics is a diffusion process. Moreover, as this dynamical system evolves, we find that the
time derivative of the expected value of parameters is proportional to the gradient of entropy
with reciprocal (symmetric) values — similar to the Onsager’s approach to nonequilibrium
statistical physics. This result is further commented in [6].

Chapter 6 presents an investigation done in collaboration with F. X. Costa and published in [4]. In it, we apply EntDyn to a system of random networks. The resulting
dynamical process solves a debate on the nature of scale-free networks. One of the best
known examples of dynamics of networks involves preferential attachment [65], leading to a
power law behaviour in node connectivity. However, such models have been criticized for not
providing a strong fit to networks found in real world data [66]. This discussion presented
the need for further dynamical models of networks, for which the power law behaviour in
scale free networks is observed only in idealized conditions. Based on previous work involving
Gibbs distributions of random graphs [34, 35, 67, 68, 69, 70, 37], we developed an EntDyn
of networks. In particular conditions, we find that the steady-state solution of our EntDyn
model fits the node degree distributions found in real world networks [66].
CHAPTER 2

Background

In this chapter, we will review some well-known background results that will be fundamental to several topics in the present thesis. Mainly we will present in section 2.1 the view of probability theory as a method to extend logic to reasoning with incomplete information. Section 2.2 will present a method to update probability distributions when new information becomes available, leading to the foundations of entropy, and address how the topic of entropy relates to thermodynamics and statistical physics. Section 2.3 will present a method to quantify the distinguishability between probability distribution with the techniques of differential geometry, leading to a geometric structure for which the metric is given by the Fisher-Rao information metric (FRIM) and the field known as information geometry.

2.1 Probabilities

This section will develop a theory of probabilities as a method for reasoning with incomplete information inspired by Cox [14] but follows arguments similar to Jaynes [71], Caticha [72] and N.Caticha [73]. This is equivalent to designing quantitative measures of “rational belief”. We choose to proceed this way instead of an axiomatic theory – such as the one by Kolmogorov found in most books – since these theories work on a robust mathematical formalism and attach ad-hoc interpretation to it. Instead, here we will describe this measure of rational belief and design the appropriate tools for dealing with them

\footnote{This comes at the cost that some mathematical proofs when applying the theory of probabilities to continuous sets will lack mathematical rigor, as we will address later.}

A final definition to what rationality means is beyond the scope of the present thesis, since these intersect with the greatest questions in philosophy. Rather, here we will make assumptions on what a quantitative rational theory must be and, through the process of eliminative induction, all theories that do not fit into these assumptions are deemed irrational. The result is that the only “non-irrational” theory left is what is known as probability theory.
2.1.1 Boolean Algebra notation

In order to proceed finding the logic of incomplete information, it is important to establish the general algebra of logic or Boolean algebra. This endeavor will start by defining a set of possible assertions all \{a, b, c, \ldots\}, namely the universe of discourse. These assertions are well-defined logical sentences for which one can attribute a truth value (that means, say it is true or false) as, for example, “Earth’s mass is between $5.95 \times 10^{24}\text{kg}$ and $6.00 \times 10^{24}\text{kg}$” or “There is life on Mars”. One must not confuse the fact that an assertion is well-defined to whether we have knowledge to determine its truth value.

Primary properties of this universe of discourse are: (i) For every assertion $a$, the universe of discourse includes the negation of this assertion denoted $\bar{a}$ which if true when $a$ is false and vice-versa. (ii) For every pair of assertions $a$ and $b$, the universe of discourse includes their conjunction denoted $a \land b$ which is true when both $a$ and $b$ are true and false otherwise. (iii) For every pair of assertions $a$ and $b$, the universe of discourse includes their disjunction denoted $a \lor b$ which is true when either $a$ or $b$ are true and false when both $a$ and $b$ are false.\footnote{Here we explained the existence of disjunction as an independent assumption. However one could derive the existence of disjunction from our other two assumptions, namely the existence of negation and conjunction. The disjunction is the negation of a conjunction of the negation of two assertions, meaning $a \lor b = \bar{c}$, where $c = \bar{a} \land \bar{b}$.} This defines what is known as the Boolean algebra, important properties implied by conjunction and disjunction are

\begin{align*}
\text{redundancy,} & \quad a \lor a = a \land a = a, \\
\text{commutivity,} & \quad a \lor b = b \lor a, \\
& \quad a \land b = b \land a, \\
\text{associativity,} & \quad a \land (b \land c) = (a \land b) \land c, \\
& \quad a \lor (b \lor c) = (a \lor b) \lor c, \\
\text{distributivity,} & \quad a \land (b \lor c) = (a \land b) \lor (a \land c), \\
& \quad a \lor (b \land c) = (a \lor b) \land (a \lor c). 
\end{align*}

(2.1)

which we will use later.
2.1.2 Designing a rational theory

Our first assumption on a quantitative theory of rational belief is that we only assign a degree of belief within some sort of context (or model). To express $a$ in terms of another assertion within the same universe of discourse $i$, we define the notation $a|i$, read as “$a$ given $i$” or “$a$ conditioned on $i$”. This is important because one cannot expect to assign a truth value to an assertion on a blank slate. Certain assertions, such as “elves exist”, have a different truth value whether one talks about the real world or some role-playing game campaign setting. Also an assertion such as “elves exist” has a completely different meaning in the context of J.R.R. Tolkien’s Lord of the rings series than in most Christmas movies, even though they share their truth value. Other assertions, such as “Darth Vader is Anakin Skywalker”, are ill-defined in the real world but are well-defined in the context of the Star Wars movie franchise.

Our second assumption on a quantitative theory of rational belief is that we can define a transitive order (or ranking) of beliefs. That means, it must be included in the theory a way to say that a assertion $a|i$ is more believable (sound, likely, plausible or, as we will determine later, probable) than another assertion $b|i$. This leads to a need for a well-ordered set of plausibility values, if $a|i$ is more believable than $b|i$ and $b|i$ is more believable than $c|i$ it must follow that $a|i$ is more believable than $c|i$. To implement this, therefore, degrees of belief must be real numbers. We will represent the degree of belief in $a|b$ by a real-valued function $f(a|b)$ – later we will associate $f$ with a monotonic function through function composition leading to the probability $p(a|b)$.

Two extreme values for $f$ will be defined. If an assertion is known to be false given its context, $f$ will map it into a specific values $v_F$, e.g. $f(\overline{a}|a) = v_F$, and for similar reasons a assertion that is true in its context is taken by $f$ into a singular value $v_T$, e.g. $f(a|a) = v_T$. It follows from the order of rational beliefs that there is a single value for $v_F$ and $v_T$.

Our third assumption on a quantitative theory of rational belief is that the degree of belief in a conjunction or disjunction of two assertions must follow directly from the degree of belief on these assertions. That means, in quantitative terms, that there are functions $F$
and $G$ such that for every pair of assertions $a$ and $b$ we have

$$f(a \lor b|c) = F(f(a|c), f(b|c), f(a|b \land c), f(b|a \land c)) ,$$

(2.2)

and

$$f(a \land b|c) = G(f(a|c), f(b|c), f(a|b \land c), f(b|a \land c)) .$$

(2.3)

$F$ is the function that gives the degree of belief in the disjunction, while $G$ gives the degree of degree in the conjunction. To clarify the notation used in (2.2) and (2.3), for the remainder of this section we must know that $a \lor b|c$ means $(a \lor b)|c$ and analogously $a|b \land c$ means $a|(b \land c)$. Later we will use different techniques to find the form of $F$ and $G$. Note that the arguments for $F$ and $G$ include all degrees of belief that can be constructed from $a$ and $b$ under the context defined by $c$. The following subsection will find the form of $F$ and $G$ based on principles of rationality.

2.1.3 Why probabilities?

Here we will explore how the usual sum and product rules of probabilities arise from the form of $F$ and $G$.

The first step to derive the form of $F$ and $G$ is to show that every disjunction can be written as a disjunction of two mutually exclusive assertions, meaning for every $a$ and $b$ there is a pair of assertions $a'$ and $b'$ so that $a \lor b = a' \lor b'$ and $a' \land b'$ is false. This follows directly if we write

$$a = (a \land b) \lor (a \land \bar{b}), \quad \text{and} \quad b = (b \land a) \lor (b \land \bar{a}),$$

(2.4)

the conjunction of $a$ and $b$ can be rewritten with the redundancy and associative properties in (2.1) as

$$a \lor b = ((a \land b) \lor (a \land \bar{b})) \lor (b \land \bar{a}) = a \lor (b \land \bar{a}) .$$

(2.5)

Let $a' = a$ and $b' = b \land \bar{a}$ therefore finishing the proof.
2.1.3.1 Disjunction and sum rule

To obtain the disjunction function, we rewrite (2.2) for mutually exclusive assertions we have

\[ f(a \lor b|c) = f((a' \lor b'|c) = F(f(a'|c), f(b'|c), v_F, v_F) \equiv F(f(a'|c), f(b'|c)) . \quad (2.6) \]

Since there are two ways to calculate the same quantity, they must agree. Therefore, it is enough to obtain \( F \) – a function of two real parameters – rather than \( F \) – with four parameters. To check its consistency with the associative property in (2.1) we write

\[ F(F(f(a'|d), f(b'|d)), f(c'|d)) = F(f(a'|d), F(f(b'|d), f(c'|d))) . \quad (2.7) \]

where \( a', b' \) and \( c' \) are mutually exclusive assertions. Since this has to hold for all possible assertions therefore for arbitrary real numbers \( x, y \) and \( z \) we have

\[ F(F(x, y), z) = F(x, F(y, z)) . \quad (2.8) \]

We can check by direct substitution that a possible solution for (2.8) is

\[ F(x, y) = \phi^{-1}(\phi(x) + \phi(y) + \kappa) , \quad (2.9) \]

where \( \phi \) is an arbitrary bijective and monotonic real valued function and \( \kappa \) is an arbitrary real constant. One can find a proof that (2.9) is a general solution for (2.8) in chapter I.3 of [14] – which is revisited in appendix B of [71] and chapter 2.2.4 of [21] – such proof is beyond the scope of the present thesis. Note that \( f \), and consequentially \( \phi \), were designed to give a real valued ranking to assertions. In order to facilitate notation we define the following function

\[ \xi(a|b) \equiv \phi(f(a|b)) + \kappa , \quad (2.10) \]

assigning again each assertion to a real value. Since \( \phi \) is monotonic, \( \xi \) preserves the order
defined by \( f \). If we substitute (2.10) and (2.9) into (2.7) we have
\[
\xi(a \lor b|d) = \xi(a|d) + \xi(b \land \bar{a}|d) \\
= \xi(a|d) + \xi(b \land \bar{a}|d) + \xi(b \land a|d) - \xi(b \land a|d).
\] (2.11)

If we use, per (2.1), that \( b \land a \) and \( b \land \bar{a} \) are mutually exclusive and \( (b \land a) \lor (b \land \bar{a}) = b \) we have
\[
\xi(a \lor b|d) = \xi(a|d) + \xi(b|d) - \xi(b \land a|d).
\] (2.12)

This equation must also follow in the case of total disbelief, if we have \( d = \bar{a} \land \bar{b} \) we obtain
\[
\xi(a \lor b|\bar{a} \land \bar{b}) = \xi(a|\bar{a} \land \bar{b}) + \xi(b \land a|\bar{a} \land \bar{b}).
\] (2.13)

Therefore, the \( \xi \) value for complete uncertainty, \( \xi_F \equiv \phi(v_F) + \kappa = \phi(f(a|\bar{a}) + \kappa \), is
\[
\xi_F = 2\xi_F \quad \rightarrow \quad \xi_F = 0.
\] (2.14)

2.1.3.2 Conjunction and product rule

A similar simplification can be done for the conjunction function in (2.3). In chapter I.3 of [14] – see also chapter 2.3.1 of [21] – it is shown how \( G \) can be rewritten, in the regraduation given by \( \phi \), as a function of two arguments instead of four meaning,
\[
\xi(a \land b|c) = \mathcal{G}(\xi(a|c), \xi(b|a \land c)).
\] (2.15)

If we apply this for three assertions \( a \), \( b \), and \( c \) in the context of another assertion \( d \) the distributive property guarantees that
\[
\xi(a \land (b \lor c')|d) = \xi(a \land b|d) + \xi(a \land c')|d)
\] (2.16)

where \( c' = c \land \bar{b} \), so that \( b \) and \( c' \) are mutually exclusive. Using (2.12) and (2.15) this can be rewritten as
\[
\xi(a \land (b \lor c')|d) = \mathcal{G}(\xi(a|d), \xi(b|a \land d) + \xi(c'|a \land d))
\mathcal{G}(\xi(a|d), \xi(b|a \land d)) + \mathcal{G}(\xi(a|d), \xi(c|a \land d)).
\] (2.17)
Note that this was obtained solely from the fact that two different ways to calculate the same quantity must agree.

In terms of real parameters we can obtain the following functional equation for $G$

$$G(u, v + w) = G(u, v) + G(u, w) .$$

(2.18)

That means, given a first parameter, $G$ is linear in the second parameter, therefore its general solution is of the form

$$G(u, v) = \alpha(u)v + \beta(u) ,$$

(2.19)

where $\alpha$ and $\beta$ are two real valued functions. Direct substitution in (2.18) yields $\beta(u) = 0$. To find $\alpha(u)$ we can write (2.15) for $b = c$, meaning

$$
\xi(a|c) = \xi(a \land c|c) = G(\xi(a|c), \xi(c|a \land c)) = \alpha(\xi(a|c))\xi_T ,
\xi_T = \phi(\nu_T) + \kappa,
$$

(2.20)

where $\xi_T = \phi(\nu_T) + \kappa$, the last equation holds as $c|a \land c = c|c$ per redundancy. Therefore

$$
\alpha(u) = \frac{u}{\xi_T} \quad \text{and} \quad G(u, v) = \frac{uv}{\xi_T} .
\xi_T = \phi(\nu_T) + \kappa,
$$

(2.21)

2.1.3.3 Obtaining probabilities

To conclude our discussion on the degree of belief, we can use the results in (2.12) and (2.21) to, finally, define the probability functions $p$ as

$$p(a|b) = \frac{\xi(a|b)}{\xi_T} .$$

(2.22)

From a state of complete knowledge it follows that $p(a|a) = 1$ and similarly (2.14) implies $p(\bar{a}|a) = 0$. If we rewrite (2.12) in terms of $p$ we obtain the sum rule

$$p(a \lor b|c) = p(a|c) + p(b|c) - p(a \land b|c) ,$$

(2.23)

and by substituting (2.21) into (2.15) we obtain the product rule

$$p(a \land b|c) = p(a|c) p(b|a \land c) .$$

(2.24)
The degree of belief $f(a|b)$ is equivalent – in the sense of preserving order – to $\xi(a|b)$ and $p(a|b)$, the last obeys the sum (2.23) and the product (2.24) rules associated to probabilities. That can be stated as “probabilities are degrees of rational belief”, certainly other representations for rational belief are possible, we will proceed using only probabilities because of its convenience.

2.1.4 Other discussions in probability theory

Having established the probability theory in this subsection, we will discuss other topics in probability theory, the concepts defined here are commonly in probability and statistics and we will refer to it often in the present thesis.

Mainly we will lay the conditions for which the sum and product rules becomes simple sums and products of probabilities. Then we will explain the process of marginalization, on which the probability of an assertion can be obtained from the probabilities of some conjunctions. We will then comment on how to assign probabilities to numerical values, as opposed to assertions. Finally, we comment on expected values and their meaning.

2.1.4.1 Independence and mutual exclusivity

The sum and product rules for probabilities can take useful forms in specific contexts. Mainly the turn into simple sums and products of probabilities for mutually exclusive and independent assertions.

Two statements are said to be independent if the knowledge (or assumption) of one does not change the degree of belief for the other. That means, a statement a is independent of b (in the context of c) if

$$p(a|b \land c) = p(a|c) ,$$

(2.25)

if we apply the product rule (2.24) and symmetry we have that

$$p(a \land b|c) = p(a|c) \ p(b|a \land c) = p(b|c) \ p(a|b \land c) ,$$

(2.26)

hence a independent of b implies b is also independent of a

$$p(b|a \land c) = p(b|c) .$$

(2.27)
For independent statements (2.24) becomes just the product of the probabilities for each statement

\[ p(a \land b|c) = p(a|c) \ p(b|c) \ . \tag{2.28} \]

In section 2.1.3 we suggest that two assertions are mutually exclusive when their conjunction is false. Holding this as a definition, if \( a \) and \( b \) are mutually exclusive the sum rule (2.23) becomes the sum of probabilities for each statement

\[ p(a \lor b|c) = p(a|c) + p(b|c) \ , \tag{2.29} \]

since \( p(a \land b|c) = 0 \). If we have a set of mutually exclusive assertions \( \{a_1, a_2, \ldots, a_n\} \), for which \( a_i \land a_j \) is false, it follows directly that

\[ p(a_1 \lor a_2 \lor \ldots \lor a_n|c) = \sum_{i=1}^{n} p(a_i|c) \ . \tag{2.30} \]

### 2.1.4.2 Marginalization

A set of assertions, \( \{a_1, a_2, \ldots, a_n\} \), is said to be exhaustive if the conjunction of all of them is necessarily true, \( a_1 \lor a_2 \lor \ldots \lor a_n = T \), therefore for any \( c \) we have

\[ p(a_1 \lor a_2 \lor \ldots \lor a_n|c) = 1 \ . \tag{2.31} \]

That can also be understood as at least one of them is true or that an exhaustive set include all possibilities. From a set of exhaustive assertions it is always possible, as explained in 2.1.3, to create a set of assertions \( \{a'_1, a'_2, \ldots, a'_n\} \) that is both exhaustive and mutually exclusive. It implies that

\[ p(a'_1 \lor a'_2 \lor \ldots \lor a'_n) = \sum_{i} p(a'_i) = 1 \ , \tag{2.32} \]

above the conditional probability is omitted, although implied, since it is valid for any conditional \( b \),

\[ p(a'_1 \lor a'_2 \lor \ldots \lor a'_n|b) = \sum_{i} p(a'_i|b) = 1 \ . \tag{2.33} \]
It follows naturally then that,

\[ p((a'_1 \lor a'_2 \lor \ldots \lor a'_n) \land b) = \sum_i p(a'_i \land b) = p(b), \tag{2.34} \]

which is known as marginalization.

Expressions like (2.34) are useful to determine the numerical values of a probability. If one has a model on which they can calculate the probability for conjunctions \( a'_i \land b \) the probability \( p(b) \) follows directly. The same argument is valid in case one can calculate the probability for \( b \) conditioned on one of the exhaustive assertions, \( p(b|a'_i) \) as

\[ p(b) = \sum_i p(b|a'_i)p(a'_i). \tag{2.35} \]

### 2.1.4.3 Random variables

Often when one talks about probabilities one is interested in data from experiments, for which it is necessary to attribute numerical values as the assertions on the universe of discourse. This is done by attributing functions \( x(a) \), namely random variables, where \( x(a) \in \mathcal{X} \) and \( \mathcal{X} \) is an algebraic number field. For example, if the universe of discourse are the results of a roll of two 6-sided dice, one can define \( x \) as the sum of the two rolled values; if the universe of discourse is the result of the tosses of \( n \)-coins, one can define \( x \) as the number of heads tossed. If \( \mathcal{X} \) is discrete, one can define the probability of the random variables as

\[ p(x) = \sum_{a \in A_x} p(a), \quad \text{where} \quad A_x = \{ a | x(a) = x \}, \tag{2.36} \]

and for a set of numbers \( B \) one can define the probability for the set as

\[ p(B) = \sum_{x \in B} p(x). \tag{2.37} \]

A physicist might find it natural to extend (2.37) to a continuous field \( \mathcal{X} \) as

\[ p(B) = \int_B dx \ p(x), \tag{2.38} \]
where \( \int dx \) is the integral measure of your space \( \mathcal{X} \). A consequence of this is that the probability of a single value in a continuous set is zero, for example if \( \mathcal{X} \) is a real number we have

\[
p(\{x\}) = \int_x^x dx' \; p(x') = 0 . \tag{2.39}
\]

Still, the generalization is correct, even though the Coxian formalism does not allow for rigorous verification. We presented the probabilities so far as the “degree of rational belief” therefore meaning is immediately attached and consistency lead to the sum (2.23) and product (2.24) rules. However in order to explain how probabilities for random variables relate to integrals we will refer to Kolmogorov’s approach – presented in most probability textbooks e.g. [74] – which defines probabilities as measures but does not give the insight on the meaning. In the present thesis we proceed using (2.38) despite the lack of mathematical rigor.

### 2.1.4.4 Expected values

If one knows the probability distribution \( p(x|c) \) defined over an algebraic field – with a defined meaning of summation – we may ask for a reasonable way to estimate the value of \( x \) within the context given by \( c \). It seems reasonable that higher values of \( p \) should contribute more for the estimate. We define the expected values of \( x, \langle x \rangle \), as

\[
\langle x \rangle = \int_{\mathcal{X}} dx \; x \; p(x|c) . \tag{2.40}
\]

Naturally one should not believe that all information for \( x \) is contained in \( \langle x \rangle[5] \). The term “expected” may be misleading, in an algebraic structure that is not complete, the expected value might not be reasonably expected. For example in a fair 6-sided die would yield the expected value 3.5 which is not an element (hence is not a result) of the universe of discourse.

The definition of expected value can be extended to any function \( f \) defined over \( x \) as

\[
\langle f \rangle = \int dx \; f(x) \; p(x|c) , \tag{2.41}
\]

---

4For an interesting discussion of the role of mathematical rigor in science see Appendix B in [71].

5This notation may sometimes leave the information codified by the conditional \( c \) unclear. When necessary we may specify the context by writing the expected value in (2.40) as \( \langle x \rangle|_c \).
note that this definition for $\langle f \rangle$ is a direct consequence of (2.40) for a smooth $f$. If we write the probability terms of the values of $f$ we have

$$\int_{\mathcal{X}_0} \mathrm{d}x \ p(x|c) = \int_{f(\mathcal{X}_0)} \mathrm{d}f \ p(f|c)$$

(2.42)

where $\mathcal{X}_0$ is an arbitrary subset of $\mathcal{X}$, $\mathcal{X}_0 \subset \mathcal{X}$, and $f(\mathcal{X}_0)$ refers to the image of $\mathcal{X}_0$ through $f$. In order for (2.42) to be valid for every $\mathcal{X}_0$ we must have:

$$p(f|c) = \left( \frac{df}{dx} \right)^{-1} p(x|c) ,$$

(2.43)

then the definition of $\langle f \rangle$ follows as

$$\langle f \rangle = \int \mathrm{d}f \ f \ p(f|c) = \int \mathrm{d}x \ \frac{df}{dx} f(x) \left( \frac{df}{dx} \right)^{-1} p(x|c) ,$$

(2.44)

matching (2.41).

The question arising is whether $\langle x \rangle$ is a good estimator of $x$. Two other estimators are also commonly defined in the statistical literature, each with their own usefulness, namely the mode $x_M$

$$x_M = \arg \max p(x|c) ,$$

(2.45)

and if $\mathcal{X}$ is well-ordered, the median $x_m$,

$$x_m : \quad p(\{x \leq x_m\}|c) = p(\{x \geq x_m\}|c) .$$

(2.46)

The mode gives the most likely value, meaning the value that is most reasonably expected to occur in the context given by $c$. While the median is the value for which the probability of being larger or smaller is equal. Note that the expected value in (2.40) takes into account every possible value in $\mathcal{X}$.

A convenient measure to define whether $\langle x \rangle$ is a good estimate of $x$ is the variance, $\text{Var}(x)$ defined as

$$\text{Var}(x) = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 .$$

(2.47)
A way to interpret variance, and its relationship to expected values, involves supposing an arbitrary estimator \( x' \) and minimizing a function that gives a difference between \( x \) and \( x' \). For example, if we want to minimize the expected value of \((x' - x)^2\) we have

\[
\frac{d}{dx'} \langle (x - x')^2 \rangle = 0 \tag{2.48}
\]

which is solved for \( x' = \langle x \rangle \), hence the expected value is the ‘best’ estimator when we define best as the one that minimizes quadratic differences.

To finish our background study/discussion on probability theory, it is also useful to define the moments of \( x \), \( m_k(x) \) as

\[
m_k(x) \doteq \langle x^k \rangle , \tag{2.49}
\]

where \( k \) is a natural number, \( m_k(x) \) reads a “the \( k \)-th moment of \( x \)”\(^6\). We can also define the “\( k \)-th” central moment of \( x \) as

\[
M_k(x) \doteq \langle (x - \langle x \rangle)^k \rangle , \tag{2.50}
\]

note that \( \text{Var}(x) = M_2(x) \).

Having established some background and usual notation in probability theory, as a theory for reasoning with incomplete information, we will comment on how to assign and update probabilities given certain information.

### 2.2 Entropy

The study of reversible heat engines by Carnot and Clausius lead to the mathematical definition of a thermal quantity labeled “entropy” for which the laws of thermodynamics were defined upon. The first major reinterpretation to the concept of entropy came from the work of Boltzmann for which entropy was understood as the logarithm of the number of accessible states at a fixed energy. Leading to the works of Gibbs, which was the first to

---

\(^6\)Equivalently to expected values we may write the \( k \)-th moment of \( x \) as \( m_k(x)|_c \) to specify the conditional \( c \).
write entropy in the form
\[ S[p] = - \int dx \ p(x) \log p(x) , \] (2.51)
reducing to Boltzmann entropy to uniform distributions.

From a completely different perspective, Shannon [12] arrived at the same expression (although for discrete states) by investigating communication theory. Even though Shannon coined the use of the word entropy in general probability theory, he did not believe that the quantity was directly related to thermodynamics. It took the work of Jaynes [15, 16, 76] to “fill the gap” by designing the method of maximum entropy and prove that, indeed, the entropy of Gibbs in thermodynamics is the Shannon entropy constrained to the conserved quantities for Hamiltonian dynamics.

Another major reinterpretation comes from the work of Kullback and Leibler [77] introducing entropy as a relative concept. This led Shore and Johnson [78] to the modern interpretation: entropy is a tool for updating the probability distribution when new information arrives. We will review the foundations of entropy based on a recent axiomatic structure [20] and discuss how the exponential family arises from it.

### 2.2.1 Caticha’s design criteria

The goal of this section is to design a mathematical tool for choosing the appropriate probability distribution over a space of possible states \( x \in \mathcal{X} \) of a system. On the understanding explained in the previous section, \( p(x)\) encodes the incomplete information on the systems’ state. MaxEnt – as designed by Jaynes [15, 71] – selects distributions that maximize entropy \( S[p] \), a constructed mathematical functional meant to represent the uncertainty – or amount of missing information – in \( p(x) \). That means the value of \( S[p] \) is how much information one would need to completely identify the state \( x \) when the system is being described by \( p(x) \). Through this design, entropy gives only the amount, it does not give what kind of information is missing or how to obtain such missing information.

This idea is inspired by a crucial concept in intellectual honesty and parsimony: we should state our assumptions and not introduce into the thinking model anything that is not

---

7There is a claim – without direct citation – in page 81 of [75] that Shannon didn’t mean to call the quantity entropy, the term was just coined due to von Neumann’s advice “no one understands entropy very well, so in any discussion you will be in a position of advantage!”.
an explicit consequence of these assumptions; while updating only on the minimum amount required by the data. This is referred to \cite{19, 20, 22} as “principle of minimum updating” (PMU): when new information becomes available, we should change – or update – probability distributions only to the extent \textit{required} by those pieces of information.

On the idea that entropy is supposed to give an amount of missing information, one would expect $S[p]$ to be large for broad distributions and small for sharp distributions. So, thinking of MaxEnt in very naive terms might be misleading. It may seem that the method always selects the broader distribution possible, ignoring all information previously known.

This leads to one of the first comments on how to apply the method: to realise that there is no such a thing as an empty statement. Because of this, in the Shore and Johnson understanding of MaxEnt, the process will update from a prior $q(x)$ meant to represent the previously known information and then search for the posterior $\rho(x)$ that maximizes entropy $-\rho = \arg \max S[p]$ – given a set of constraints representing the available information at hand. If new information is to be added, we would search for an entropy maximum solely among probability distributions that are in accordance with the new piece of information.

So, to proceed into finding the functional form of $S[p]$ we need to understand that entropy has to be a functional of both $p$ and the prior $q$, written as $S[p|q]$, as inspired by the notation of conditional probability. Note that, as sometimes we omit what a probability distribution is conditional upon, we will eventually equally write entropy as just $S[p]$ omitting the prior in which such entropy is “conditioned”. A functional that depends both on the prior and the posterior is usually stated as a relative entropy. Although such nomenclature may be useful, it is important to know that fundamentally every entropy is relative.

The design criteria (DC) coined by Caticha \cite{19, 21, 22, 8} summarize basic principles of statistical inference. These criteria can pinpoint the form of $S[p|q]$ by elimination. Out of all functionals possible, only the ones who are in accordance to the DC are allowed. The only functional that appropriately validates both is the KL entropy.

**DC1 - Subdomain independence**: in a space of states $\mathcal{X}$ partitioned into complementary subdomains spaces, $\mathcal{X} = \mathcal{D}_1 \cup \mathcal{D}_2$ and $\mathcal{D}_1 \cap \mathcal{D}_2 = 0$. If information about the

\footnote{We will talk about two DC, namely subdomain independence and subsystem independence. The first work presenting these DC \cite{19} also stated a third criterion, coordinate invariance, which has been found not to be necessary for uniquely defining the form of entropy \cite{20, 22}. It is relevant to point out that the elimination of such criterion leads to the entropic formalism for density matrices \cite{20}.}
states of only one of the subsystems – say $D_1$ – becomes available, the conditional probability for elements of the other subsystem should not be updated. That means, the posterior probability that maximizes entropy must be so that

$$\rho(x|D_2) = q(x|D_2), \quad (2.52)$$

where $\rho = \arg \max S[p]$, the posterior probability assigned by MaxEnt.

**DC2 - Subsystem independence**: in a space of states $X$ can be taken as a composition of two spaces, $x = (x_1, x_2) \in X = X_1 \times X_2$, note that in such a space a probability $p(x)$ would be a joint probability for the state of two subsystems, $p(x_1, x_2)$. If the prior distribution is statistically independent for the subsystem, $q(x_1, x_2) = q(x_1)q(x_2)$, and new information about the states of only one of the subsystems – say $x_1$ – becomes available, the probability for the other subsystem should not be updated. That means the posterior probability that maximizes entropy must be so that:

$$\rho(x) = \rho(x_1, x_2) = \rho(x_1)q(x_2). \quad (2.53)$$

Although both design criteria deal with subdivisions in the space of states, they do so in systematically different manners. In DC1 subdomains probabilities should be summed, $\int dx \, p(x) = \int_{D_1} dx \, p(x) + \int_{D_2} dx \, p(x) = 1$, and the probability of each subdomain is given by the full set measure, $p(D_1) = \int_{D_1} dx \, p(x)$. While in DC2 subsystems probabilities are given by composition, $\int dx \, p(x) = \int dx_1 dx_2 \, p(x_1, x_2) = 1$, and the probability of a particular substate will be given by marginalization, $p(x_1) = \int dx_2 \, p(x_1, x_2)$.

In [22] it is proven that DC1 and DC2 lead uniquely to an entropy of the form

$$S[p|q] = - \int dx \, p(x) \log \left( \frac{p(x)}{q(x)} \right). \quad (2.54)$$

The functional $S[p|q]$ is the negative Kullback-Leibler divergence [77] and $q(x)$ is a prior distribution. $S[p]$ reduces to Shannon entropy [12] when $q(x)$ is uniform.
2.2.2 Gibbs distribution/Exponential Family

The most usual application of MaxEnt consists of assigning probabilities distributions $\rho$, defined on a general space $x \in \mathcal{X}$, that maximizes entropy (2.54) under expected value constraints for a set of $n$ real functions $a^i(x)$ and normalization,

$$
\int dx \, \rho(x) = 1
$$

$$
\int dx \, \rho(x) a^i(x) = A^i .
$$

(2.55)

In order to obtain the maximum of (2.54) given the constraints defined in (2.55) one needs to solve the functional extreme condition under the Lagrange Multiplier method. Using techniques of variational calculus and functional derivatives, it leads to

$$
\frac{\delta}{\delta \rho} \left[ - \int dx \, \rho(x) \log \left( \frac{\rho(x)}{q(x)} \right) - \lambda_0 \int dx \, \rho(x) - \sum_{i=1}^{n} \lambda_i \int dx \, \rho(x) a^i(x) \right] = 0 ,
$$

(2.56)

where $\lambda_0$ is the Lagrange multiplier related to the normalization constraint, and $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ are the set of Lagrange multipliers related to the expected value constraints. After some algebra (2.56) becomes

$$
\log \left( \frac{\rho(x)}{q(x)} \right) + 1 + \lambda_0 + \sum_{i=1}^{n} \lambda_i a^i(x) = 0 .
$$

(2.57)

The distributions that achieve the entropy maximum in (2.54) given (2.55) are Gibbs distributions,

$$
\rho(x|\lambda_1, \lambda_2, \ldots, \lambda_n) = \frac{q(x)}{Z(\lambda)} \exp \left( - \sum_{i=1}^{n} \lambda_i a^i(x) \right) ,
$$

(2.58)

where $Z(\lambda)$ is a normalization factor independent of $x$, identified from (2.56) as $e^{1+\lambda_0}$, which can be obtained in terms of $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ as

$$
Z(\lambda) = \int dx \, q(x) \exp \left( - \lambda_i a^i(x) \right) = e^{-F(\lambda)} ,
$$

(2.59)

Further explanations on functional derivatives and variational calculus are beyond the scope of the present thesis. Nevertheless the discussion presented in chapter 2 of [79] and chapter 3 of [80] give a good mathematical foundation to the topic.
Above, and on the remainder of the present thesis, we use Einstein’s summation notation
\[ A_i B^i = \sum_{i=1}^{n} A_i B^i. \]

An important identification, that given the universality of MaxEnt comes at no surprise, is that Gibbs distributions are what statistics literature refers to as the exponential family. The exponential family is the only form of distributions for which sufficient statistics exist, meaning when sampling no other statistics need to be calculated in order to give the complete form of the distribution \(^{10}\). Incidentally, the functions \( a^i(x) \) for expected value constraints are taken in MaxEnt are the sufficient statistics.

The expected values \( \langle a^i(x) \rangle = A^i \) are written with upper indexes as the coordinates of the manifold. For that reason we write \( a^i(x) \) with upper indexes (contravariant) and \( \lambda_i \) by lower indexes (covariant).

At its maximum (2.58), the entropy (2.55) can be computed as a function (rather than a functional) of the expected values,
\[ S(A) = - \int dx \, \rho(x|\lambda(A)) \log \frac{\rho(x|\lambda(A))}{q(x)} = \lambda_i(A)A^i - F(\lambda(A)) . \]
which is the macrostate entropy. Equation (2.60) shows how \( S(A) \) is the Legendre transform of \( F(\lambda) \). Usually, in statistical physics, the macrostate entropy \( S(A) \) is what we call the thermodynamic entropy \(^{76}\) or – when there is no room for confusion – just entropy. Note that \( F(\lambda) \) plays the role of the thermodynamic free energy per temperature. Note that from (2.59) one obtains
\[ \frac{\partial F}{\partial \lambda_i} = - \frac{1}{Z(\lambda)} \frac{\partial Z(\lambda)}{\partial \lambda_i} = \int dx \, \rho(x|\lambda) a^i(x) = A^i . \]
Similarly, from (2.60) one obtains
\[ \frac{\partial S}{\partial A^i} = \frac{\partial \lambda_\alpha}{\partial A^i} A^\alpha + \lambda_i - \frac{\partial F}{\partial \lambda_\alpha} \frac{\partial \lambda_\alpha}{\partial A^i} = \lambda_i , \]
where (2.61) was used in the last step.

\(^{10}\)Interestingly, this was a problem proposed by Fisher \(^{38}\) in the primordium of statistics and later proved independently by Pitman \(^{40}\), Darmois \(^{39}\), and Koopman \(^{41}\).
distributions might be too restrictive. However, under different choices of sufficient statistics $a(x)$ and priors $q(x)$, many known probability distributions can be written as \[ (2.58) \] \[ [81, 82] \]. This is, in itself, a success of MaxEnt as a fundamental principle of inference. A short summary on how to obtain some known distributions is presented at \textbf{Table 2.1}.

\subsection*{2.2.3 MaxEnt and thermodynamics}

In this section we are going to talk about results of MaxEnt that are particular to physics. Mainly, we are going to discuss Jaynes’ proof of the second law of thermodynamics from MaxEnt \[ [76] \], and how extensivity arises (or not) from MaxEnt. Energy, Hamiltonians and other physics centered concepts will be of interest here, but they are not fundamental in MaxEnt. The choice of constraints in statistical physics is justified in terms of thermal baths.

\subsubsection*{2.2.3.1 Second law of thermodynamics}

For the purpose of this subsection, a physical system is defined by saying that the state $x \in \mathcal{X}$ (or microstate) follows a Hamiltonian dynamics. Meaning the states are the phase space for $N$ particles $- x = \Pi_{j=1}^{N} (\varphi_j, \phi_j)$ where $\varphi_j$ and $\phi_j$ are, respectively, the generalized coordinate and momentum for the particle labeled $j$ and $q(x)$ is uniform\footnote{Similar arguments to the ones presented here justify the uniformity of $q(x)$ see e.g. \[ [21] \]. In that sense the result of this section, \[ (2.68) \], should be held as a principle – a deterministic dynamics involves no loss of information implying a uniform prior.} and a probability distribution that evolves with it, $p(x,t)$ follows the Liouville equation

\begin{equation}
\frac{\partial p(x,t)}{\partial t} = \{H, p\}, \tag{2.63}
\end{equation}

where $\{A, B\}$ is the Poisson bracket. It is well known that \[ (2.63) \] also implies a local continuity equation

\begin{equation}
\frac{\partial p(x,t)}{\partial t} = - \sum_{i=1}^{N} \frac{\partial}{\partial \phi_i} \left( p(x,t) \frac{d\phi_i}{dt} \right) + \frac{\partial}{\partial \varphi_i} \left( p(x,t) \frac{d\varphi_i}{dt} \right). \tag{2.64}
\end{equation}

Understanding the timely evolution of $p(x,t)$, we can calculate the evolution of its entropy, meaning we define the function
| Distribution          | \( \rho(x|a, \beta) \) | \( \lambda \) | Suff. Stat. | Prior |
|-----------------------|-------------------------|-----------|-------------|-------|
| Exponent Polynomial   | \( \frac{\beta}{\beta^2} \sqrt{\frac{\beta}{\pi}} e^{-\beta x} \) | \( \lambda = \beta \) | \( a(x) = x \) | uniform |
| Gaussian              | \( \frac{\beta}{\pi \sigma^2} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2} \) | \( \lambda = -\log \sigma \) | \( a(x) = (x, x) \) | uniform |
| Poisson               | \( \frac{\mu}{\mu^2} x! e^{-\mu} \) | \( \lambda = -\log \mu \) | \( a(x) = (x) \) | uniform |
| Multinomial (k)       | \( \frac{n!}{x_1! \cdots x_k!} \theta_1^{x_1} \cdots \theta_k^{x_k} \) | \( \lambda = -\log(\theta_1, \theta_2, \ldots, \theta_k) \) | \( a(x) = (x_1, \ldots, x_k) \) | uniform |
| Mixed power laws      | \( x^{\alpha-1} e^{-\beta x} \) | \( \lambda = (\alpha, \beta) \) | \( a(x) = (\log x, x) \) | uniform |
we can calculate
\[
\frac{dS(t)}{dt} = -\int dx \frac{\partial}{\partial t} (p(x,t) \log p(x,t)) = -\int dx \left( \frac{\partial p(x,t)}{\partial t} \log p(x,t) \right),
\]
which can be substituted using (2.64) and calculated by partial integration and taking into account that \( p(x,t) \) is a probability as
\[
\frac{dS(t)}{dt} = \int dx \left( \sum_{i=1}^{N} \frac{\partial}{\partial \phi_i} \left( p(x,t) \frac{d\phi_i}{dt} \right) + \frac{\partial}{\partial \varphi_i} \left( p(x,t) \frac{d\varphi_i}{dt} \right) \right) \log p(x,t) = \int dx \left( \frac{\partial p(x,t)}{\partial t} \right) = 0.
\]

Where the second line uses partial integration and the third line uses Liouville equation (2.63). And, therefore, we have
\[
\frac{dS(t)}{dt} = 0.
\]
That means, the entropy (2.65) for a probability evolving in a Hamiltonian dynamics does not change in time.

Interestingly, this argument could have been done in the opposite order. Understanding entropy as the amount of missing information, one could notice that (2.68) holds in a Hamiltonian dynamics, meaning that deterministic dynamics should lead to no loss of information. The only way to guarantee it is with a functional entropy of the form (2.54) and uniform prior.

However, a similar result would not hold for the sufficient statistics. In a time interval from \( t \) to \( t' > t \) the entropy of the distribution \( S \) would not change, \( S(t) = S(t') \). Each sufficient statistics \( a_i(x) \) would have it's average values as a function of time defined by
\[
A_i(t) = \int dx \, a_i(x)p(x,t) .
\]
and, in general \( A_i(t) \) is not equal to \( A_i(t') \). Therefore if the system starts at an equilib-
rium distribution, meaning $p(x,t) = p(x|A)$ and therefore $S(t) = S(A(t))$ and reaches an equilibrium at a time $t'$, we know from (2.68) that $S(t) = S(t')$ 

$$S(A(t)) = S(t) = S(t') \leq S(A(t')).$$

(2.70)

Which is the main result of [76]. The consequence is that, in a Hamiltonian dynamics, between two states of equilibrium the macrostate entropy (2.60) evolves only to bigger values. Hence the second law of thermodynamics is valid for $S(A)$.

As one can notice, this discussion about the second law of thermodynamics involves not only the functional entropy $S[p]$ but also two functions derived from it $S(A)$ and $S(t)$. All of them are fairly called entropies in the information theory sense, however they represent very different statistical objects. It is the macrostate entropy $S(A)$ that appropriately takes the role of the thermodynamic entropy, meaning it is the one that follows the second law of thermodynamics. Still, that proof relies directly on the fact that it comes from the KL entropy functional $S[p|q]$ in (2.54). The derivation shows that a similar macrostate entropy defined from a different functional would not be compatible with the second law of thermodynamics.

2.2.3.2 Extensivity

A property of KL entropy, which is in itself a consequence of DC2, is its additivity. That means, when subsystems are statistically independent, both in the prior and posterior, $q(x) = q(x_1)q(x_2)$ and $p(x) = p(x_1)p(x_2)$, the entropy of the composite system, namely joint entropy, is the sum of the entropies for each subsystem,

$$S[p] = -\int dx \, p(x) \log \frac{p(x)}{q(x)} = -\int dx_1 \, p(x_1) \log \frac{p(x_1)}{q(x_1)} - \int dx_2 \, p(x_2) \log \frac{p(x_2)}{q(x_2)}.$$

(2.71)

A proper way to express additivity is to use a notation in which entropy is subscripted by the set in which the probability distribution is defined, meaning for $y \in \mathcal{Y}$

$$S_y = S[p] = -\int dy \, \rho(y) \log \frac{\rho(y)}{q(y)},$$

(2.72)

allowing (2.71) to be written as $S_{X_1 \times X_2} = S_{X_1} + S_{X_2}$.
Thermodynamics deal with large systems that can be divided into several self-similar smaller subsystems, \( X = X_1 \times X_2 \times \ldots \times X_N \). At this point, we have no assumptions on what these subsystems are. If such subdivision is statistically independent apriori, \( q(x) = \prod_{j=1}^{N} q(x_j) \), we can examine a case in which each sufficient statistics can be written as a sum of these statistics for each subsystem,

\[
a_i(x) = \sum_{j=1}^{N} \tilde{a}_i(x_j),
\]

(2.73)

where \( \tilde{a}_i(x_j) \) are functions of \( x_j \in X_j \) only.

In the ideal gas model, these conditions follow directly. The subsystems can be a division particle by particle, and the sufficient statistics sum as in (2.73). Energy, for example, is just the sum of the energies of each of the \( N \) particles. We should not expect, however, that these subdivisions will always hold particle by particle. In a broader class of thermodynamical systems, one is only interested in the properties of very large systems (thermodynamic limit). In those cases, we can subdivide the system into subsystems that are small compared to the composite system, but large enough so that it can still be described in the thermodynamic limit. On this division, if all interactions are short-range, then the sufficient statistics can be approximated by (2.73). A graphical representation can be seen in figure 2.1.

In a division as this, the maximum entropy distribution (2.58) factors for each subsystem \( x_j \),

\[
\rho(x|A) = p(x_1, x_2, \ldots, x_N|A) = \prod_{j=1}^{N} \frac{q(x_j)}{Z_j(A_j)} \exp \left[ -\sum_{i} \lambda_i \tilde{a}_i(x_j) \right].
\]

(2.74)

This maximum entropy distribution is the same as if we were to apply separately for each subsystem \( j \), updating from the subsystem prior factor \( q(x_j) \) under the constraint,

\[
\int dx_j \ p(x_j) \tilde{a}_i(x_j) = \tilde{A}_{i,j},
\]

(2.75)

that is a subsystem equivalent of the constraints in (2.55). The posterior distribution \( \rho(x_j|\tilde{A}_j) \) will be

\[
\rho(x_j|\tilde{A}_j) = \frac{q(x_j)}{Z_j(A_j)} \exp \left[ -\sum_{i} \lambda_{i,j} \tilde{a}_i(x_j) \right].
\]

(2.76)

where each \( \lambda_{i,j} \) is the Lagrange multiplier associated to \( A_{i,j} \). As these subsystems are small
compared to the composite system, if allowed to interact they would yield the same Lagrange multiplier $\lambda_{i,j} = \lambda_i$, $\forall j$, as explained in [15]. This allows (2.74) to be written as

$$\rho(x|A) = \prod_{j=1}^{N} \rho(x_j|\tilde{A}_j).$$

That means, as expected by DC2, the maximum entropy for a priori independent composite subsystem, under constraints that are separable, keeps the subsystems statistically independent.

We identify Gibbs distributions, between constraints in (2.55) and (2.73), it follows that $A_i = \sum_j \tilde{A}_{i,j}$. Also, the thermodynamic entropy for these subsystems, $S(\tilde{A}_j)$, can be computed by noticing that the entropy functional is additive and it is at its maximum. Applying this in the condition where the subsystems are identical – so that each function $\tilde{a}_i(x)$ and expected value $\tilde{A}_{i,j}$ are the same for all $j$ – we obtain

$$S(A) = S(A_1, A_2, \ldots A_n) = N S(A_1/N, A_2/N, \ldots A_n/N),$$

that means thermodynamic entropy is a homogeneous function of degree 1, therefore extensive.

The extensivity, as a property of a system’s thermodynamic entropy, is a consequence of (2.73), unlike additivity (2.71) which is a general property of the entropy functional, independent on what kind of system it is applied on. This is a source of great confusion. In a non-statistical understanding of thermodynamics, extensivity of entropy is taken as a fundamental postulate [42], that may lead people [83] to be surprised if a system happens to not have extensive entropy. That should be understood as an emergent property of a system that cannot be approximated by (2.73), not as a failure of entropic methods. Systems that contain long-range interactions (plasmas, black holes) or with considerable surface effects are not expected, under the additive entropy functional, to have extensive thermodynamic entropy.
Figure 2.1: A graphical representation of the conditions for extensivity. On a system composed of several subsystems each sufficient statistic for the composite system (in the gas example energy, volume and number of particles for each molecule species) will be approximately the sum of a analogous sufficient statistics calculated at each subsystem. In physics this will be violated, for example, when the boundary energy or long-range interactions are not negligible even in the thermodynamic limit.

2.3 Information geometry

Having established a meaning for probabilities – a tool for reasoning with incomplete information – and a meaning to entropy – the tool to assign and update probabilities when new information arrives – one might be lead to ask whenever one would, instead, pick a similar distribution. How can we define the similarity between two distributions? And what errors would come from it? This raises the need for a measure of distinguishability between probabilities, which will be addressed in the present section.

In order to address distinguishability, one needs to parametrize a family of probability distributions, meaning choosing a countable set of parameters $\theta = \{\theta^i\}$ and defining a subset of probability distributions of the form $p(x|\theta)$. This is a map between $\theta$ and $p(x|\theta)$, which we will call a manifold by attributing a Riemannian metric structure, meaning defining distances $d\ell$ between neighbouring distributions – in the map – $p(x|\theta)$ and $p(x|\theta + d\theta)$ as

$$d\ell^2 = g_{ij}d\theta^i d\theta^j,$$

where $g_{ij}$ is called the Riemannian metric and $\theta$ are the coordinates of the manifold.
The field known as information geometry (IG) consists of assigning the Riemannian metric as the Fisher-Rao information metric\(^\text{12}\) (FRIM), meaning

\[ g_{ij} = \int dx \, p(x|\theta) \frac{\partial \log p(x|\theta)}{\partial \theta^i} \frac{\partial \log p(x|\theta)}{\partial \theta^j}. \] (2.80)

In this section, we will present some general concepts of Riemannian geometry, followed by derivations of (2.80) in different contexts. Finally, we will discuss some useful identities raising from calculating FRIM in (2.80) for Gibbs distributions defined in (2.2.2). Concluding that probabilities have an intrinsic notion of distances.

### 2.3.1 Basic definitions in Riemannian geometry and information geometry

Two important concepts in differential and Riemannian geometry, namely volume and curvature, will be often commented upon later in the present thesis. These are defined in this subsection. Our goal here is not to give a lengthy motivation with the necessary detail to understand how Riemannian geometry works, but rather only to give the fundamental properties of these concepts in order to guarantee the self-containment of the present thesis, as these definitions will be important in the following chapters.

#### 2.3.1.1 Length and metric

First we must say that coordinates \(\theta\) in (2.79) are, in principle, completely arbitrary. When one uses a differently defined set of coordinates \(\tilde{\theta}^i = f^i(\theta^1, \theta^2, \ldots)\). In Riemannian geometry, it is necessary to have distances as an invariant, meaning as coordinates transform from \(\theta\) to \(\tilde{\theta}\), the metric transforms from \(g_{ij}\) to \(\tilde{g}_{ij}\) so that the distance between two points along a curve – represented by \(\theta(\tau)\) for \(\tau \in (0, 1)\) does not change, therefore

\[ \ell = \int d\ell = \int_0^1 d\tau \left[ g_{ij} \frac{\partial \theta^i}{\partial \tau} \frac{\partial \theta^j}{\partial \tau} \right]^{1/2} = \int_0^1 d\tau \left[ \tilde{g}_{ij} \frac{\partial \tilde{\theta}^i}{\partial \tau} \frac{\partial \tilde{\theta}^j}{\partial \tau} \right]^{1/2}. \] (2.81)

In order to guarantee that this holds in general, it follows that

\[ \tilde{g}_{mn} = g_{ij} \frac{\partial \theta^i}{\partial \theta^m} \frac{\partial \theta^j}{\partial \theta^n}, \] (2.82)

\[^\text{12}\text{It must be understood that Fisher is considered the father of statistics and he indeed wrote (2.80) as a matrix in [84]. It was only after the work of Rao [85] that (2.80) was defined as a Riemannian metric.}\]
that means the metric transforms covariantly\textsuperscript{13}.

2.3.1.2 Volume element

One important quantity in differential geometry, whose influence will be further discussed here, is the concept of volumes in a Riemannian manifold. In a Euclidean space with rectangular coordinates $\theta$, one expects that a compact set $\mathcal{A}$ will have its volume $V$ defined by the integral

$$V(\mathcal{A}) = \int_{\mathcal{A}} \prod_i d\theta^i,$$

(2.83)

or alternatively, the volume element around a point to be of the form $dV = \prod_i d\theta$.

As done with lengths, one can identify what the extended notion of volume element by observing how it should behave under a change of coordinates $\tilde{\theta} = f(\theta)$. An invariant volume implies an integral for which the value of the integrand is invariant over any set, meaning we have to find a function $\mathcal{V}$ so that for any $\mathcal{A}$ we have

$$\int_{\mathcal{A}} \prod_i d\theta^i \mathcal{V}(\theta) = \int_{\mathcal{A}} \prod_m d\tilde{\theta}^m \mathcal{V}(\tilde{\theta}).$$

(2.84)

Note that from (2.83) it follows that in a Euclidean space one must have $\mathcal{V}(\tilde{\theta}) = 1$.

Well-known results from integration in several variables lead us to rewrite (2.84) as

$$\int_{\mathcal{A}} \prod_i d\theta^i \mathcal{V}(\theta) = \int_{\mathcal{A}} \prod_m d\tilde{\theta}^m \mathcal{V}(\tilde{\theta}) \cdot F \mathcal{V}(\tilde{\theta}),$$

(2.85)

where

$$F = \det F^m_i, \quad \text{with} \quad F^m_i = \frac{\partial \tilde{\theta}^m}{\partial \theta^i}.$$

(2.86)

Since $\mathcal{A}$ is an arbitrary set, it follows that $\mathcal{V}(\theta) = F \mathcal{V}(\tilde{\theta})$. The calculation of $F$ as a function $\mathcal{V}(\theta)$ following this discussion.

\textsuperscript{13}Although further comment on this topic is slightly beyond the scope of the present thesis it is useful to mention the definition of covariant and contravariant transformations. A quantity $s$ is said to be a true scalar when, in a change of coordinates $\tilde{\theta} = f(\theta)$, $s(\theta) = s(\tilde{\theta})$. A quantity $V^i$ transforms contravariantly when $V^i(\theta) = \frac{\partial \tilde{\theta}^m}{\partial \theta^i} V^m(\tilde{\theta})$. Analogously, we say that a quantity $W^i$ transforms covariantly when $W^i(\theta) = \frac{\partial \tilde{\theta}^m}{\partial \theta^i} W_m(\tilde{\theta})$. This usually justifies the use of upper and lower indexes through the present thesis. From (2.82) we can see that the metric $g_{ij}$ transforms twice covariantly or 2-covariantly – as a product of two quantities that each transform covariantly. In the present thesis, we will use a broader definition, whenever a mathematical object transforms accordingly to the positions of its indexes, we will say it transforms covariantly.
of the metric terms can be done by substituting the definition of $F_i^m$ in (2.86) into (2.82) and taking the determinant, yielding

$$\tilde{g} = g \, F^{-2},$$

(2.87)

where $g = \det g_{ij}$.

In the particular case where the original coordinates $\theta$ are rectangular in a Euclidean space, it follows that $F = \tilde{g}^{-1/2}$. In general, if we adopt $V$ to be of the form $V(\theta) = \sqrt{\tilde{g}}$ calculated at $\theta$, it is straightforward to see that it observes (2.85) and, consequentially, (2.84) for an arbitrary set $A$. Hence, the square root of the metric determinant is the appropriate generalized volume element, $dV = \prod_i d\theta^i \sqrt{\tilde{g}}$.

2.3.1.3 Christoffel symbols and covariant derivative

One important definition in differential geometry is the concept of Christoffel symbols, which are defined at a point $\theta$ of the manifold as

$$\Gamma^i_{jk} = \frac{1}{2} g^{i\alpha} \left( \frac{\partial g_{k\alpha}}{\partial \theta^j} + \frac{\partial g_{j\alpha}}{\partial \theta^k} - \frac{\partial g_{jk}}{\partial \theta^\alpha} \right).$$

(2.88)

The importance of this object is tied to how derivatives of covariant and contravariant objects behave under a change of coordinates. Let $V$ be a contravariant object in a change of coordinates from $\theta$ to $\tilde{\theta}$, we directly have $V^i = \frac{\partial \theta^i}{\partial \tilde{\theta}^m} \tilde{V}^m$ and, therefore, if one takes a derivative $A$ to some parameter $\tau$, it leads to

$$\frac{dV^i}{d\tau} = \frac{\partial \theta^i}{\partial \tilde{\theta}^m} \frac{d\tilde{V}^m}{d\tau} + \tilde{V}^m \frac{\partial^2 \theta^i}{\partial \tilde{\theta}^m \partial \tilde{\theta}^n} \frac{d\tilde{\theta}^n}{d\tau}.$$  

(2.89)

Hence, derivatives of a contravariant object do not generally transform as a contravariant object. Considering this, a question arises on whether it is possible to obtain a derivative of a covariant vector that by itself transforms covariantly.

It is relevant to also notice how the Christoffel symbols defined in (2.88) transform on a change of coordinates, which one can calculate directly as

$$\Gamma^i_{jk} = \frac{\partial \theta^i}{\partial \tilde{\theta}^m} \frac{\partial \tilde{\theta}^n}{\partial \theta^j} \frac{\partial \tilde{\theta}^p}{\partial \theta^k} \Gamma^m_{np} + \frac{\partial^2 \tilde{\theta}^m}{\partial \theta^i \partial \theta^k} \frac{\partial \theta^j}{\partial \tilde{\theta}^m}.$$  

(2.90)
Therefore Christoffel symbols also do not transform covariantly. One can notice, however, that both $\frac{dV^i}{d\theta^j}$ and $V^\alpha \Gamma^{i}_{j\alpha}$ have their deviation from transforming covariantly by the same term.

This leads one to define the covariant derivative of a contravariant object as

$$D_i V_j = \frac{\partial V_j}{\partial \theta^i} + \Gamma^j_{i\alpha} V^\alpha,$$  \hfill (2.91)

and, similarly, the covariant derivative of a covariant object is defined as

$$D_i V_j = \frac{\partial V_j}{\partial \theta^i} - \Gamma^\alpha_{ij} V^\alpha.$$  \hfill (2.92)

These do transform covariantly under a change of coordinates, meaning

$$D_i V_j = \frac{\partial \tilde{V}_j}{\partial \tilde{\theta}^i} + \tilde{\Gamma}^j_{i\alpha} \tilde{V}^\alpha, \quad \text{and} \quad D_i V_j = \frac{\partial \tilde{V}_j}{\partial \tilde{\theta}^i} - \tilde{\Gamma}^\alpha_{ij} \tilde{V}^\alpha.$$  \hfill (2.93)

Further discussion on the topic can be seen in any text on differential geometry or general relativity – e.g. [86, 87, 88, 89, 90] – where covariant derivatives are used to define the concept of parallel transport – a method to transport vectors along curves in a non-Euclidean space – nevertheless, such discussion is beyond the scope of the present thesis.

It is relevant to present a few interesting identities relating to Christoffel symbols – which will be found relevant in the following chapters. First, if one contracts the indexes on the Christoffel symbols, meaning $\Gamma^i_{ik}$, one obtains from (2.88)

$$\Gamma^i_{ik} = \frac{1}{2} g^{i\alpha} \frac{\partial g_{i\alpha}}{\partial \theta^k}.$$  \hfill (2.94)

In parallel, one can observe that

$$\frac{\partial g}{\partial \theta^k} = g g^{\alpha\beta} \frac{\partial g_{\alpha\beta}}{\partial \theta^k},$$  \hfill (2.95)

therefore one can write (2.94) as

$$\Gamma^i_{ik} = \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g}}{\partial \theta^k}.$$  \hfill (2.96)
A similar identity can be found for covariant divergences $\mathcal{D}_i V^i$

$$\mathcal{D}_i V^i = \frac{\partial V^i}{\partial \theta^i} + \Gamma^i_{ia} V^a = \frac{\partial V^i}{\partial \theta^i} + \frac{1}{\sqrt{g}} \frac{\partial \sqrt{g}}{\partial \theta^i} V^i = \frac{1}{\sqrt{g}} \frac{\partial}{\partial \theta^i} \sqrt{g} V^i . \tag{2.97}$$

2.3.1.4 Curvature

Given the metric tensor of a Riemannian manifold, $g_{ij}$, and obtaining the Christoffel symbols $\Gamma^i_{jk}$, one can move towards a measure of how “curved” a space is. Here, the definitions that answer this question will be presented. For a discussion on how these definitions arise from extensions of two dimensional surfaces in a three dimensional space see [86]. The role of curvature in information geometry is a matter of heated debate – see e.g. [50, 91, 92]. The present thesis will contribute to the topic by discussing the curvature obtained from the statistics of quantum gases in Chapter 3.

The first definition towards this goal is the Riemann curvature tensor, defined as

$$R^i_{jkl} = \frac{\partial}{\partial \theta^k} \Gamma^i_{jl} - \frac{\partial}{\partial \theta^l} \Gamma^i_{jk} + \Gamma^i_{ka} \Gamma^a_{jl} - \Gamma^i_{la} \Gamma^a_{kl} . \tag{2.98}$$

When all the elements of the Riemann curvature tensor are zero, $R^i_{jkl} = 0$, the manifold is termed a flat manifold. A few important identities follow directly from this definition. First, if one reduces the upper index in (2.98) – obtaining a fully covariant tensor $R_{ijkl} = g_{ia} R^a_{ijkl}$ – one obtains

$$R_{ijkl} = \frac{1}{2} \left( \frac{\partial^2 g_{it}}{\partial \theta^i \partial \theta^t} - \frac{\partial^2 g_{jt}}{\partial \theta^j \partial \theta^t} - \frac{\partial^2 g_{ik}}{\partial \theta^i \partial \theta^l} + \frac{\partial^2 g_{jk}}{\partial \theta^j \partial \theta^l} \right) + g_{a\beta} \left( \Gamma^\alpha_{jk} \Gamma^\beta_{il} - \Gamma^\alpha_{ja} \Gamma^\beta_{kl} \right) . \tag{2.99}$$

In this form, some symmetries in the curvature tensor are straightforward. Namely, the curvature tensor is antisymmetric in the first pair and last pair of indexes

$$R_{ijkl} = R_{jikl} \quad \text{and} \quad R_{ijkl} = R_{ijlk} , \tag{2.100}$$

and it is symmetric in the exchange between the first and last pair of indexes,

$$R_{ijkl} = R_{kl ij} . \tag{2.101}$$
From these symmetries it follows that

\[ R_{ijkl} + R_{iljk} + R_{iklj} = 0 , \]  
\[ (2.102) \]

which is known as the first Bianchi identity. Differentiating \((2.98)\) it is also straightforward to obtain

\[ \frac{\partial}{\partial \theta^m} R_{ijkl} + \frac{\partial}{\partial \theta^k} R_{ijkm} + \frac{\partial}{\partial \theta^l} R_{ijkm} = 0 , \]  
\[ (2.103) \]

known as the second Bianchi identity.

Note that the terms of the Riemann curvature tensor are coordinate dependent. One might imagine a coordinate independent (scalar) for the curvature. In order to do so, it is useful to define the Ricci tensor, \( R_{ij} \)

\[ R_{ij} = R^\alpha_{i\alpha j} , \]  
\[ (2.104) \]

which is symmetric, \( R_{ij} = R_{ji} \). From this, one obtains the scalar curvature \( R \) as the contraction of \( R_{ij} \) with the metric tensor

\[ R = g^{\alpha \beta} R_{\alpha \beta} = g^{\alpha \beta} g^{\mu \nu} R_{\alpha \beta \mu \nu} , \]  
\[ (2.105) \]

as the coordinate independent measurement of curvature.

### 2.3.2 Derivations of FRIM

In this section, we will discuss how the FRIM arises from geometric considerations on a space of probability distributions. Two derivations will be presented: the first tries to use the relative difference as the distinguishability between infinitesimally close probability distributions, while the second extends the relative entropy between neighbouring probability distributions to a continuous path in the space of parameters. Although both arrive at the metric form in \((2.80)\), the final argument for the fact that such metric is unique is in the work of Čencov \[93\] – for modern discussion see also \[94, 95\] – which proves that \((2.80)\) is the only metric invariant under Markovian embeddings.
2.3.2.1 FRIM from relative difference

The first attempt to create a measure of distinguishability for probability distributions is the intuitive discussion by Rao for which the distinguishability between two neighbouring probability distributions is given by the relative difference

\[ \Delta = \frac{p(x|\theta + d\theta) - p(x|\theta)}{p(x|\theta)}, \]  

(2.106)

which, considering the Taylor’s expansion given by

\[ \Delta = \frac{\partial \log p(x|\theta)}{\partial \theta^i} d\theta^i + o(\theta^i), \]  

(2.107)

the differential will be understood here by ignoring terms smaller than \( d\theta \). Then the expected value for the relative difference is given by

\[ \langle \Delta \rangle = \int d\mathbf{x} p(x|\theta) \Delta = d\theta^i \int d\mathbf{x} p(x|\theta) \frac{\partial \log p(x|\theta)}{\partial \theta^i}, \]  

(2.108)

That means, the expected value of the relative difference vanishes, which is not surprising given the definition in (2.106). However, the variance of \( \Delta \) does not vanish, rather

\[ \text{Var}(\Delta) = \langle \Delta^2 \rangle = \int d\mathbf{x} p(x|\theta) \frac{\partial \log p(x|\theta)}{\partial \theta^i} \frac{\partial \log p(x|\theta)}{\partial \theta^j} d\theta^id\theta^j. \]  

(2.109)

We will use the variance of relative difference as our quadratic difference between two neighbouring distributions \( d\ell^2 = \langle \Delta^2 \rangle \) substituting (2.109) into the definition of Riemannian metric (2.79) yields

\[ g_{ij} = \int d\mathbf{x} p(x|\theta) \frac{\partial \log p(x|\theta)}{\partial \theta^i} \frac{\partial \log p(x|\theta)}{\partial \theta^j}, \]  

(2.110)

as in (2.80).

The recognition of \( \langle \Delta^2 \rangle \) as a distance and \( g_{ij} \) as a metric is an important contribution of Rao [85] and the founding stone of IG. Interestingly, we will proceed into other forms of arriving at the same metric.
2.3.2.2 FRIM from Relative entropy

As discussed in section 2.2, the concept of entropy is a tool for updating a probability distribution \( q \) to another distribution \( \rho \) that adopts the new information made available about the system, as \( \rho \) maximizes the entropy \( S[\rho|q] \). This may lead to the idea that one might consider that the updated distribution \( \rho \) is the “closest” to \( q \) that apply the constraints described by the new information. The first issue with this line of thought is that the relative entropy (2.54) between \( \rho \) and \( q \) is not symmetric – meaning \( S[\rho|q] \neq S[q|\rho] \). Because of this the relative entropy could not be understood as a distance.

Although entropy cannot be directly identified with any notion of distance (or inverse distance), one can continue the investigation on connecting probability, entropy and geometry by understanding entropy as a non-local concept and define the distance between infinitesimally close distributions. Following the ideas of differential geometry, this is implemented by studying the relative entropy between two probability distributions indexes by a set of parameters \( \theta \) and a neighbouring distribution \( \theta + d\theta \), a direct Taylor expansion of entropy around \( \theta \) gives

\[
S_{\theta}(\theta + d\theta) = S_{\theta}(\theta) + \frac{\partial S_{\theta}(\theta')}{\partial \theta^n} \bigg|_{\theta' = \theta} d\theta^i + \frac{1}{2} \frac{\partial^2 S_{\theta}(\theta')}{\partial \theta^n \partial \theta^j} \bigg|_{\theta' = \theta} d\theta^i d\theta^j + o(d\theta^2),
\]

(2.111)

where \( S_{\theta}(\theta') = S[p(x|\theta')|p(x|\theta)] \). Note that \( S_{\theta}(\theta) = 0 \) and it is straightforward to see from Gibbs inequality that \( S_{\theta} \leq 0 \). Therefore, \( S_{\theta} \) has a global maximum at \( \theta \) and is concave, meaning \( \frac{\partial^2 S_{\theta}}{\partial \theta^n \partial \theta^j} \leq 0 \). Hence (2.111) becomes

\[
S_{\theta}(\theta + d\theta) = \frac{1}{2} \frac{\partial^2 S(\theta')}{\partial \theta^n \partial \theta^j} \bigg|_{\theta' = \theta} d\theta^i d\theta^j.
\]

(2.112)

when expanded up to the second order in \( d\theta \).

As done in section 2.3.2 one could identify \( S_{\theta} \) as a negative quadratic distance, meaning \( \frac{1}{2} d\ell^2 = -S_{\theta}(\theta + d\theta) \) from which one can identify

\[
g_{ij} = -\frac{\partial^2 S(\theta')}{\partial \theta^n \partial \theta^j} \bigg|_{\theta' = \theta} = \int dx \ p(x|\theta) \frac{\partial \log p(x|\theta)}{\partial \theta^i} \frac{\partial \log p(x|\theta)}{\partial \theta^j},
\]

(2.113)

37
2.3.3 Information geometry of Gibbs distributions

This section will study some useful identities for FRIM when it is calculated for Gibbs distributions of the form (2.58). Alternatively one can interpret this as the metric when one writes the probability in the Gibbs form – as presented in Table 2.1. Gibbs distributions generalize a broad range of well known probability distributions.

The canonical probability distributions (2.58) were defined conditional to the values of $\lambda$. One might argue that the appropriate way to define the information metric is to take the derivatives related to the values of $\lambda$ instead of the values of $A$. However, the difference between doing the information geometry in terms of $A$ instead of $\lambda$ is not more than a change of variables, that can be transformed using

$$\frac{\partial A^i}{\partial \lambda_k} = -\frac{\partial^2 \log Z}{\partial \lambda_k \partial \lambda_i} = A^k A^i - \langle a^k a^i \rangle,$$

(2.114)

where we identify the covariance tensor and its inverse as

$$C_{ij} = \langle a^i a^j \rangle - A^i A^j = -\frac{\partial A^i}{\partial \lambda_j}, \quad C_{ij} = -\frac{\partial \lambda_i}{\partial A^j}. \quad (2.115)$$

Applying (2.115) into the calculation of the metric tensor (2.80) one obtains

$$g_{ij} = \int dx \rho(x|A) C_{ik} \frac{\partial \log \rho(x|A)}{\partial \lambda_k} C_{jl} \frac{\partial \log \rho(x|A)}{\partial \lambda_l} , \quad (2.116)$$

so that we can calculate

$$\frac{\partial \log \rho(x|A)}{\partial \lambda_k} = \frac{\partial}{\partial \lambda_k} \left(-\lambda_i a^i(x)\right) - \frac{\partial \log Z}{\partial \lambda_k} = A^k - a^k(x). \quad (2.117)$$

Therefore, the metric tensor is a covariance matrix,

$$g_{ij} = C_{ik} C_{jl} C^{kl} = C_{ij}, \quad (2.118)$$

38
and the indexes for the small change in the coordinates \( dA^i \) can be lowered as

\[
dA_i = g_{ij} dA^j = C_{ij} dA^j = -\frac{\partial \lambda_i}{\partial A^j} dA^j = -d\lambda_i.
\] (2.119)

That means, the covector \( dA_i \), for a small change \( dA^i \) is the negative of vector for small changes in \( d\lambda_i \). Not only these coordinates have a dual relation through the Legendre transformation \([2.60]\) – which is a consequence of maximization of entropy – but also their vectors are also dual in the statistical manifold\(^{14}\).

The FRIM for Gibbs distributions \([2.118]\) can also be conveniently written in terms of \( S(A) \) as

\[
g_{ij}^A = -\frac{\partial^2 S}{\partial A^i \partial A^j}. \tag{2.120}
\]

Note that \([2.120]\) only refers to Gibbs distributions in the coordinate system defined by the expected values \( A \) as coordinates. If one wants to find the metric in another system of coordinates, namely \( \theta^i(A) \) the metric transforms as

\[
g_{ij} = \frac{\partial A^\alpha}{\partial \theta^i} \frac{\partial A^\beta}{\partial \theta^j} g_{ij}^A = -\frac{\partial A^\alpha}{\partial \theta^i} \frac{\partial A^\beta}{\partial \theta^j} \frac{\partial^2 S}{\partial A^\alpha \partial A^\beta}. \tag{2.121}
\]

In the particular case where these coordinates are the dual Lagrange multipliers \( d\lambda^i = g_{ij}^A dA^j \lambda^i \) – it is straightforward to notice that

\[
g_{ij}^\lambda = -\frac{\partial^2 F}{\partial \lambda^i \partial \lambda^j}. \tag{2.122}
\]

In chapter 3 – where I discuss the geometry of Gibbs distributions related to Fermi-Dirac and Bose-Einstein statistics – the free-energy and Lagrange parameters form of the information metric \([2.122]\) is preferred. In Chapters 5 and 6 – where a more general branch of Gibbs distributions is studied – the macrostate entropy and expected values form of the information metric \([2.120]\) will be favored.

### 2.3.3.1 Curvature of two dimensional Gibbs manifolds

It is interesting to notice that per \([2.122]\) and \([2.120]\) that, in the coordinates given by the Gibbs distributions, the metric is given as the Hessian of a function. If the metric of

\(^{14}\)This also justifies the use of lower indices for \( \lambda_i \) in section \([2.2.2]\).
a two dimensional system happens to be given as a Hessian of a function $T$ with respect to its coordinates, meaning
\[ g_{ij} = \frac{\partial^2 T(\theta)}{\partial \theta^i \partial \theta^j}, \]  
(2.123)
it follows that the scalar curvature can be written as
\[ R = -\frac{1}{2g^2} \det \begin{bmatrix} g_{11} & g_{12} & g_{22} \\ \partial_1 g_{11} & \partial_1 g_{12} & \partial_1 g_{22} \\ \partial_2 g_{11} & \partial_2 g_{12} & \partial_2 g_{22} \end{bmatrix} \]  
(2.124)
where $\partial_\sigma g_{\mu\nu} = \partial^\alpha g_{\mu\nu}$.

Verifying (2.124) is relatively straightforward. In such system of coordinates the Christoffel symbols are written by substituting (2.123) into (2.88), which leads to
\[ \Gamma^i_{jk} = \frac{1}{2} g^{i\alpha} \frac{\partial^3 T}{\partial \theta^\alpha \partial \theta^j \partial \theta^k}, \]  
(2.125)
which one can substitute in (2.99) obtaining
\[ R_{ijkl} = g_{\mu\nu} \left( \Gamma^\mu_{jk} \Gamma^\nu_{il} - \Gamma^\mu_{jl} \Gamma^\nu_{ik} \right) 
\quad = \frac{1}{4} g^{\alpha\beta} \left( \frac{\partial^3 T}{\partial \theta^\alpha \partial \theta^j \partial \theta^k} \frac{\partial^3 T}{\partial \theta^\beta \partial \theta^i \partial \theta^l} - \frac{\partial^3 T}{\partial \theta^\alpha \partial \theta^i \partial \theta^l} \frac{\partial^3 T}{\partial \theta^\beta \partial \theta^j \partial \theta^k} \right) \]  
(2.126)
due to the symmetries in the $R_{ijkl}$ – presented in section 2.3.1.4 – the only non-vanishing components in a two-dimensional manifold are $R_{1212} = R_{2121} = -R_{2112} = -R_{1221}$. The scalar curvature is given by (2.105), which in two dimensions and using (2.126) is equivalent to
\[ R = (g^{11} g^{22} - g^{12} g^{21}) R_{1212} = \frac{1}{2 \det g_{\mu\nu}} \det \begin{bmatrix} g^{22} & g^{12} & g^{11} \\ \partial_1 g_{11} & \partial_1 g_{12} & \partial_1 g_{22} \\ \partial_2 g_{11} & \partial_2 g_{12} & \partial_2 g_{22} \end{bmatrix}. \]  
(2.127)
Knowing the terms for the inverse of a two dimensional matrix $g^{11} = \frac{g_{22}}{\det g_{\mu\nu}}$, $g^{12} = -\frac{g_{21}}{\det g_{\mu\nu}}$, $g^{21} = -\frac{g_{22}}{\det g_{\mu\nu}}$, and $g^{22} = \frac{g_{11}}{\det g_{\mu\nu}}$ – one can notice that (2.127) is equivalent to (2.124).
CHAPTER 3
Information geometry for Fermi-Dirac and Bose-Einstein quantum statistics

3.1 Introduction

Information geometry (IG) is the application of differential geometry to probability theory [96, 97, 98]. The geometric structure of the space of probability distributions represents the distinguishability between neighbouring probability distributions. The metric that achieves this is the Fisher-Rao [84, 85] information metric (FRIM) which is the only metric which is invariant under Markov embeddings [94]. This finds application in a large number of information science disciplines including machine learning [99], signal processing [100] as well as quantum information [101] and statistical physics. IG has also been found to derive quantum mechanics in an entropic formalism [63, 102].

Weinhold was one of the first scientists to apply elements of geometry into study of the structure of equilibrium statistical mechanics [103] as an extension of Einstein’s fluctuation theory and the axioms of thermodynamics. It was Ruppeiner who first introduced a well-defined Riemannian metric structure as the Hessian of the thermodynamical entropy [49]. In a modern understanding, this geometrical structure in statistical physics is due to Jaynes’ description of thermodynamics from information theory [15, 16], in which Gibbs distributions arise from maximization of entropy (MaxEnt) under expected value constraints. The geometric structure found by Ruppeiner is equivalent to the IG of Gibbs distributions [104, 95].

Gibbs distributions defined over the space of second quantization in quantum mechanics – usually referred to as Fock space and parametrized by the number of particles occupying each possible quantum state – are known in thermodynamics as the Fermi-Dirac (FD) statistics for fermionic particles and the Bose-Einstein (BE) statistics for bosonic particles. These distributions are important not only to provide foundations of statistical mechanics by describing the systems with quantum mechanics, but are also particularly relevant for material sciences, astrophysics and cosmology.
The Riemannian geometric formulation of thermodynamics and statistical mechanics has opened up a wide range of applications extended to investigations concerning complexity, phase transitions, and critical phenomena [105, 52, 106]. A key new tool introduced is Riemannian scalar curvature which yields relevant information about interparticle interactions in the physical system under investigation. The original result that motivated such a physical interpretation of the thermodynamic curvature was the fact that in an ideal gas there is no effective interparticle interaction and, remarkably, the scalar curvature of the manifold that characterizes a classical ideal gas is identically zero [49]. This significance of curvature was due to Ruppeiner, according to whom the curvature is positive when attractive interactions dominate and is negative when repulsive interactions are dominant [51].

Expanding on that, we point out that Ruppeiner proposed [49] the so-called “interaction hypothesis” according to which the scalar curvature in the context of thermodynamics geometry can be viewed as a measure of interaction in gas systems [49, 50]. Specifically, if interparticle interactions between particles are not present, as in an ideal gas, the space of thermodynamic states is flat. In the presence of interparticle interactions, the space is curved and exhibits nonzero curvature. In particular, curvature has units of volume and diverges to infinity near a critical point. Despite this very appealing interpretation, as Ruppeiner himself pointed out [51], there is no explicit quantitative connection between scalar curvature and phase transitions. In the sign convention according to which the scalar curvature of a two-sphere of radius $a$ is positive (that is, $R = 2/a^2$), it happens that [46, 47, 48, 107]: i) $R = 0$ for an ideal classical gas; ii) $R < 0$ for an ideal FD gas; iii) $R > 0$ for an ideal BE gas. As pointed out by Brody and Rivier [104], there is no definitive physical interpretation of the scalar curvature in the context of information geometric investigations of statistical mechanics of gases. To the best of our knowledge there is no mathematical proof that the scalar curvature diverges at critical points. This divergent behavior remains a conjecture to be proved or disproved by means of a suitable counterexample [104].

Moreover, the meaning of the sign of the scalar curvature is also unclear. In particular, it becomes even more complicated for interacting statistical mechanics systems and, to the best of our knowledge, nowadays there is no common agreement on what the this sign means [107]. Janyszek and Mrugala [46] used the concept of quantum mechanical exchange effect to suggest a physical interpretation of the sign of the scalar curvature for ideal quantum gases.
In [46], Janyszek and Mrugala used the concept of quantum mechanical exchange effect to suggest a physical interpretation of the sign of the scalar curvature for ideal quantum gases. In the sign convention according to which the scalar curvature of a two-sphere of radius $a$ is positive (that is, $R = 2/a^2$), it happens that [46, 47, 48, 107]: i) $R = 0$ for an ideal classical gas; ii) $R < 0$ for an ideal FD gas; iii) $R > 0$ for an ideal BE gas. In the interpretation given by Janyszek and Mrugala [46] the scalar curvature is an indicator of stability of the physical system under investigation, with the understanding that a stable system exhibits thermodynamically negligible fluctuations. Even in a non interactive model, particles in a bosonic gas tend to agglomerate in the same quantum state (that is, attractive exchange effect) leading to a positive space correlations and, as a consequence, bigger fluctuations. For this reason, bosonic gases are less stable gas than an ideal classical gas and are characterized by a positive $R$. Fermions, instead, are more stable than classical particles since effectively they repeal other particles in the same quantum state (that is, repulsive exchange effect) yielding negative space correlations and, consequently, smaller fluctuations. Therefore, fermions are specified by a negative $R$.

Posterior studies on the information geometry of quantum gases appear to agree to that interpretation. For instance, Oshima and collaborators [47] support of instability interpretation proposed by Janyszek and Mrugala by performing a numerical curvature analysis of an ideal quantum gas obeying Gentile’s statistics. Similarly, Mirza and Mohammadzade [109, 48] investigate the IG properties of an ideal gas of fractional statistical particles obeying Gentile’s statistics.

Focusing on information geometric investigations of ideal quantum gases [46, 47, 48, 110], we bring into the discussion that for bosonic systems the number of particles in the ground state is unbounded and it is well-known that at low temperatures (more specifically, in the limit of the fugacity approaching one and the temperature of the gas approaching zero) a thermodynamically relevant fraction of particles reside in the ground state, which explains the phenomena known as BE condensation.

To the best of our knowledge, the information geometric investigations carried out so far in the literature concerning the BE condensation ignored the particles in the ground state.

---

\(^{15}\)For completeness, we point out that Gentile statistics [108] is a generalization of the FD and BE quantum statistics where no more than $p$ particles are allowed to occupy the same quantum state. One recovers the FD and the BE statistics when $p = 1$ and $p \to \infty$, respectively.
Such effect do not appear in FD statistics since there cannot be more than one particle in each state.

The investigation presented in [5] and the present chapter is motivated by two main reasons. First, the lack of a clear physical interpretation of scalar curvature and its sign motivates us. This lack, in particular, opens up the possibility of proposing theoretical models accommodating critical phenomena where curvature is not necessarily expected to diverge. Second, the lack of an information geometric analysis of an ideal BE gas that includes the BE condensation drives our proposed investigation proposed here.

Combining information geometry with Jaynes’ information theory approach to statistical mechanics, we seek to obtain a general formula for the FRIM for both FD and BE statistical models. In particular, we aim at investigating geometrical quantities derived from such a metric, including volume elements and scalar curvature. Our approach is unique and original for a number of reasons: (i) Unlike previous works, we calculate the above-mentioned geometrical quantities for a generic value of the density of states exponent \( \eta \), that specifies the gas model. Therefore, the formulas derived here are applicable to a broader range of quantum systems once \( \eta \) is selected; (ii) Unlike previous works, we take into account the particles in the ground state of a bosonic system. To the best of our knowledge, this is the first time IG is done considering the ground state correction for BE statistics. Therefore, our information geometric investigation will give us an appropriate description near BE condensation. For a more transparent interpretation of our results, we also present illustrative plots of suitable geometrical quantities with the help of mpmath python library [111]. In particular, this library gives a precise calculation for the polylogarithm function [112] which is fundamental for both BE and FD quantum statistics. The codes for all graphs presented here are available at my GitHub repository [61].

The layout of the rest of the chapter is as follows: In Section 3.2, we will review how the Gibbs distributions of the grand canonical ensemble are defined in Fock spaces. In Section 3.3, we will discuss the assumptions involved in the continuous approximation, focusing on how different quantum mechanics microscopic models lead to different values for the density of states exponent \( \eta \) and how the continuous approximation leads to the polylogarithm family of functions. In Section 3.4, we will calculate FRIM, volume elements and scalar curvature for an FD ideal gas and give a graphical presentation of our results. In Section 3.5, we
calculate the same geometrical quantities and graphs for a BE ideal gas and compare them to the ones calculated without the ground state corrections. In Section 3.6, we comment on the classical limit of geometrical quantities. Since both low fugacity and high temperature are required for an ideal FD or BE gas to behave as a classical ideal gas, curvature can give more about the system’s structure not available in the Legendre structure.

3.2 MaxEnt in Fock Spaces

MaxEnt requires the choice of appropriate space $\mathcal{X}$ and prior $q(x)$. For the accurate description of quantum gases, the probabilities should be assigned to the space of occupation numbers of each quantum state (Fock space). That means $x = \{x_i\}$ where $i$ corresponds to an enumeration of the eigenstates obtained from quantum mechanics, each corresponds to an eigenvalue of energy $\epsilon_i$. Each quantity $x_i$ takes the value of the number of particles on the $i$th state. Fermi-Dirac (FD) statistics refers to a state in which $x_i$ takes binary values $\{0, 1\}$, corresponding to the Pauli exclusion principle for Fermions, while in Bose-Einstein (BE) statistics $x_i$ can take natural number values $\{0, 1, \ldots\}$. These quantum statistics are usually represented in the grand canonical ensemble, that means the sufficient statistics are chosen as the energy $a_1(x) = \sum_i \epsilon_i x_i$ and the total number of particles $a_2(x) = \sum_i x_i$.

That leads to a Gibbs distribution (2.58) of the form

$$
\rho(x|\lambda) = \rho(x_1, x_2, \ldots | \lambda) = \frac{1}{Z(\lambda)} \prod_i e^{-\lambda^1 \epsilon_i x_i} e^{-\lambda^2 x_i},
$$

(3.1)

where

$$
Z(\lambda) = \sum_x \prod_i e^{-\lambda^1 \epsilon_i x_i} e^{-\lambda^2 x_i}.
$$

(3.2)

For FD and BE statistics $Z(\lambda)$ is given by

$$
Z_{FD}(\lambda) = \prod_i \left( 1 + e^{-\lambda^1 \epsilon_i} e^{-\lambda^2} \right) \quad \text{and} \quad Z_{BE}(\lambda) = \prod_i \left( 1 - e^{-\lambda^1 \epsilon_i} e^{-\lambda^2} \right)^{-1},
$$

(3.3)

respectively. The partition function and free energy (2.59) can be written as

$$
Z_{\pm}(\lambda) = \prod_i \left( 1 \pm e^{-\lambda^1 \epsilon_i} e^{-\lambda^2} \right)^{\pm 1} \quad \text{and} \quad F_{\pm}(\lambda) = \pm \sum_i \log \left( 1 \pm e^{-\lambda^1 \epsilon_i} e^{-\lambda^2} \right),
$$

(3.4)
respectively. The upper (or lower) sign in (3.4) refers to the FD (or BE) statistics. The Lagrange multipliers in (3.4) can be identified in terms of the usual thermodynamical parameters, temperature $T$ and chemical potential $\mu$,

$$\lambda^1 = \frac{1}{kT} = \beta \quad \text{and} \quad \lambda^2 = -\frac{\mu}{kT} ,$$

(3.5)

where $k$ is the Boltzmann constant. We will keep expressing geometrical quantities in terms of the Lagrange multipliers as they are the appropriate coordinates as described before.

The next step in the MaxEnt description is to write the expected values in terms of the Lagrange multipliers as in (2.61). This yields,

$$U_{\pm} = A_{1\pm} = \frac{\partial F_{\pm}}{\partial \lambda^1} = \sum_i \frac{\epsilon_i}{e^{\lambda^1 \epsilon_i} e^{\lambda^2 \pm 1}}$$

$$N_{\pm} = A_{2\pm} = \frac{\partial F_{\pm}}{\partial \lambda^2} = \sum_i \frac{1}{e^{\lambda^1 \epsilon_i} e^{\lambda^2 \pm 1}} .$$

(3.6)

If one correctly assigns the spectrum of energies $\epsilon_i$, the statistical physics description of the FD and BE gases would be complete. However to write the summations in (3.6) in a closed form is not feasible, to the best of our knowledge. The study of quantum gases is usually tackled through approximations in which the summations are replaced by integrals in energy, as we will discuss in the following section.

### 3.3 Continuous approximation

Usually in physics, we do not rely on summation such as the ones in (3.6). Instead quantum gases and other physically relevant models based on FD and BE statistics are usually treated as a continuous energy approximation,

$$\sum_i \rightarrow \int \text{d}\epsilon \ G(\epsilon) ,$$

(3.7)

where $G(\epsilon)$ is the density of states with energy $\epsilon$. This is usually justified under the idea that the internal energy of the system, $U$ is much bigger than the differences in energy of the quantum eigenstates, $\max_{ij} |\epsilon_i - \epsilon_j| \ll U$. 

Table 3.1: Parameters for density of states $G(\epsilon)$. The parameter $\kappa$ depends on $m$, the mass of the particle, and for the harmonic trapped gas depends on the frequency of the harmonic oscillator $\omega$. Table previously published in [5].

In order to better understand the continuous approximation (3.7) we will provide a first basic example. For the model of a particle in a 3 dimensional cubic box of edge length $L$, the quantum state will be defined by a triplet of integers, $i = (n_x, n_y, n_z)$ and the energy state will be given by

$$\epsilon = \epsilon_i = \frac{(2\pi \hbar)^2}{2mL^2} \left( n_x^2 + n_y^2 + n_z^2 \right),$$  (3.8)

where $m$ is the mass of the particle. Therefore the sum over states becomes

$$\sum_i = g_s \sum_{n_x=-\infty}^{\infty} \sum_{n_y=-\infty}^{\infty} \sum_{n_z=-\infty}^{\infty} \rightarrow g_s \int_{-\infty}^{\infty} dn_x \int_{-\infty}^{\infty} dn_y \int_{-\infty}^{\infty} dn_z,$$  (3.9)

where $g_s$ is the number of different possible polarizations given the particle spin $s$, $g_s = 2s+1$. The triple integral in (3.9) can be evaluated using (3.8) under spherical coordinates

$$\sum_i \rightarrow \int_0^{\infty} d\epsilon \left[ \frac{g_s V}{4\pi^2} \left( \frac{2m}{\hbar^2} \right)^{3/2} \right] \epsilon^{1/2},$$  (3.10)

where $V = L^3$ is the volume of the box.

We will proceed with the calculation of FRIM on models for which the density of states is proportional to a power of the energy,

$$G(\epsilon) = \kappa \epsilon^\eta,$$  (3.11)

where the density of states prefactor ($\kappa$) and the density-energy exponent ($\eta$) are constants defined by the quantum energy spectrum. As an example, for a three dimensional particle in the box we can see from (3.10) that $\kappa = \left[ \frac{g_s V}{4\pi^2} \left( \frac{2m}{\hbar^2} \right)^{3/2} \right]$ and $\eta = 1/2$. Other examples of
quantum systems for which (3.11) is valid and their respective values of $\kappa$ and $\eta$ are presented in Table 3.1. Note that per (3.11) and (3.7) $\kappa$ has units of $[e]^{-\eta/(\eta+1)}$, meaning that $\kappa^{-1/(\eta+1)}$ creates as an emergent unit of energy defined by the system. The continuous approximation (3.7) transforms the average values (3.6) as

$$U_\pm = A_1^\pm = \kappa \int d\epsilon \frac{e^{\eta+1}}{e^{\lambda_1 \epsilon} e^{\lambda_2 \pm 1}}$$

$$N_\pm = A_2^\pm = \kappa \int d\epsilon \frac{e^\eta}{e^{\lambda_1 \epsilon} e^{\lambda_2 \pm 1}}.$$ (3.12)

In order to proceed with the calculation of these expected values and geometric quantities derived from them, it is useful to introduce the polylogarithm function \[112\]

$$\text{Li}_\varphi(y) = \frac{1}{\Gamma(\varphi)} \int_0^\infty du \frac{u^{\varphi-1}}{y^{1-e^u - 1}} = \sum_{k=1}^\infty \frac{y^k}{k^{\varphi}},$$ (3.13)

where $\Gamma(\varphi)$ is the Euler’s gamma function. A property of the polylogarithm that is useful for obtaining geometrical quantities is

$$\frac{d}{dy} \text{Li}_\varphi(y) = \frac{1}{y} \text{Li}_{\varphi-1}(y).$$ (3.14)

The expected values (3.12) can be expressed in terms of polylogarithms. Under the appropriate change of variables, $u = \beta \epsilon$, and defining the fugacity $\xi$ as

$$\xi = e^{-\lambda^2},$$

so that, \[
\frac{d\xi}{d\lambda^2} = -\xi. \quad (3.15)
\]

Substituting (3.13) into (3.12) we obtain

$$U_\pm = A_1^\pm = \frac{\partial F}{\partial \lambda^1} = \mp \kappa \frac{\Gamma(\eta + 2)}{\beta^{\eta+2}} \text{Li}_{\eta+2}(\mp \xi)$$

$$N_\pm = A_2^\pm = \frac{\partial F}{\partial \lambda^2} = \mp \kappa \frac{\Gamma(\eta + 1)}{\beta^{\eta+1}} \text{Li}_{\eta+1}(\mp \xi).$$ (3.16)

Here we expressed the thermodynamical quantities in terms of $\beta$ defined in (3.5), and the fugacity, $\xi$ defined in (3.15). This choice is made because we commit to using the Lagrange multipliers $\lambda$ as coordinates. Moreover, we can calculate the metric terms using (2.122). If one uses a different set of coordinates one would need to express their coordinates in terms
Figure 3.1: Dimensionless quantity $\bar{g}_f$, related to metric determinant as (3.18) for fermions. Figure previously published in [5].

Figure 3.2: Dimensionless quantity $\bar{R}_f$, related to metric determinant as (3.21) for fermions. Figure previously published in [5].

of $\beta$ and $\xi$ – as we do in (3.5) and (3.15) – and also do the appropriate set of coordinates from (2.122). A similar calculation for the free energy (3.4) in the continuous approximation yields

$$F_\pm(\lambda) = \pm \kappa \frac{\Gamma(\eta + 1)}{\beta^{\eta+1}} \text{Li}_{\eta+2}(\mp \xi) . \tag{3.17}$$

Having general expressions for the expected values (3.12) allows for the calculation of FRIM for both FD and BE. Although we focus on a unified treatment for both cases, other considerations arising from the continuous approximation suggest that their treatment should be separated. Mainly the fact that (3.11) assigns no states for $\epsilon = 0$ leads to a fundamental difference in the description of bosons and fermions. In the two following sections we will calculate the geometrical quantities for the FD and BE gas, respectively.

### 3.4 Information geometry - Fermions

Applying (3.16) and (3.17) into (2.122) we can calculate the metric terms for FD models,

$$g_{11} = -\frac{\partial^2 F}{\partial \lambda^1 \partial \lambda^1} = -\frac{\partial U}{\partial \lambda^1} = -\kappa \frac{\Gamma(\eta + 3)}{\beta^{\eta+3}} \text{Li}_{\eta+2}(\mp \xi) ,$$

$$g_{12} = g_{21} = -\frac{\partial^2 F}{\partial \lambda^2 \partial \lambda^1} = -\frac{\partial N}{\partial \lambda^1} = -\kappa \frac{\Gamma(\eta + 2)}{\beta^{\eta+2}} \text{Li}_{\eta+1}(\mp \xi) , \tag{3.18}$$

$$g_{22} = -\frac{\partial^2 F}{\partial \lambda^2 \partial \lambda^2} = -\frac{\partial N}{\partial \lambda^2} = -\kappa \frac{\Gamma(\eta + 1)}{\beta^{\eta+1}} \text{Li}_\eta(\mp \xi) .$$
The derivatives were taken using (3.14) and (3.15). The metric determinant can be directly calculated from (3.18),
\[ g = \left( \frac{\kappa}{\beta n + 2} \right)^2 \bar{g}_f, \quad \text{where} \quad \bar{g}_f = A(-\xi, \eta), \] (3.19)
and \( A(x, \eta) \) is the unitless quantity defined as
\[ A(x, \eta) = \det \begin{pmatrix} \Gamma(\eta + 3)Li_{\eta+2}(x) & \Gamma(\eta + 2)Li_{\eta+1}(x) \\ \Gamma(\eta + 2)Li_{\eta+1}(x) & \Gamma(\eta + 1)Li_{\eta}(x) \end{pmatrix}, \] (3.20)
a graph for \( \bar{g}_f \) is presented in Fig. 3.1. Similarly, the scalar curvature can be calculated from (2.12) obtaining
\[ R = \frac{1}{2\bar{g}^2} \frac{\kappa^3}{\beta^{3n+7}} B(-\xi, \eta) = \frac{\beta^{n+1}}{2\kappa} \bar{R}_f, \] (3.21)
where
\[ \bar{R}_f = \frac{B(-\xi, \eta)}{A(-\xi, \eta)^2}, \] (3.22)
and \( B(x, \eta) \) is defined as
\[ B(x, \eta) = \det \begin{pmatrix} \Gamma(\eta + 3)Li_{\eta+2}(x) & \Gamma(\eta + 2)Li_{\eta+1}(x) & \Gamma(\eta + 1)Li_{\eta}(x) \\ \Gamma(\eta + 4)Li_{\eta+2}(x) & \Gamma(\eta + 3)Li_{\eta+1}(x) & \Gamma(\eta + 2)Li_{\eta}(x) \\ \Gamma(\eta + 3)Li_{\eta+1}(x) & \Gamma(\eta + 2)Li_{\eta}(x) & \Gamma(\eta + 1)Li_{\eta-1}(x) \end{pmatrix}. \] (3.23)

As discussed before, \( \kappa \) has units so that \( \frac{\beta^{n+1}}{\kappa} \) is unitless. A graph for \( \bar{R}_f \) in (3.21) is presented in Fig. 3.2. Also a visual representation of \( R \) in terms of \( \beta \) and \( \xi \) is presented in Fig. 3.3. These graphs are simple and smooth, meaning that for FD models there are no anomalies or singularities. Furthermore, it also can be seen from the Fig. 3.2 and Fig. 3.4 that the curvature for FD gases is negative.

### 3.5 Information geometry - Bosons

A similar geometric description of bosonic gases, meaning obtain FRIM from differentiating (3.17) as it was done in the previous section for fermions, would yield a metric determinant and scalar curvature given, respectively, as
Figure 3.3: Countour plot of $R$ in (3.21) as a function of $\xi$ and $\beta$ – in units of $\kappa = 1$ – for fermions in the values of $\eta = 1/2$ and $\eta = 2$. Figure previously published in [5].

\[ g_0 = \left( \frac{\kappa}{\beta^{\eta+2}} \right)^2 A(\xi, \eta) \quad \text{and} \quad R_0 = -\frac{\beta^{\eta+1}}{2\kappa} \frac{B(\xi, \eta)}{(A(\xi, \eta))^2}, \]  

(3.24)

where $A$ and $B$ are defined in (3.20) and (3.23). These are denoted as $g_0$ and $R_0$ to indicate that they do not take into account the particles in the ground state. Works such as [46, 47] are limited to this description. In this section, we will calculate FRIM and other geometrical quantities appropriately taking into account the ground state.

The ground state term correction is necessary as the continuous approximation (3.11) assigns no particles in the ground state, $G(0) = 0$. An accurate description needs to account for condensation as (3.16), and consequentially (3.24) would be an approximation for high temperatures. This is achieved by an adding the term corresponding to the number of partiles in the ground state, $\epsilon = 0$, 

\[ N_0 = \frac{1}{e^{\lambda^2} - 1}, \]  

(3.25)

in (3.12). Therefore $N$ in (3.16) for bosons becomes 

\[ N = A_2 = \kappa \frac{\Gamma(\eta + 1)}{\beta^{\eta+1}} \text{Li}_{\eta+1}(\xi) + \frac{1}{\xi^{-1} - 1}. \]  

(3.26)

and no change in the internal energy $U$ is necessary as the ground state does not contribute to the average energy.
In this correction, since all we have are the final expressions for \(U\) in (3.16) and \(N\) in (3.26), we will proceed with the metric calculation using (2.118) and (2.115), this is akin to the thermodynamical geometry of Ruppeiner [50]. This leads to the metric terms

\[
\begin{align*}
g_{11} &= -\frac{\partial U}{\partial \lambda^1} = \kappa \Gamma(\eta + 3) \text{Li}_{\eta+2}(\xi), \\
g_{12} &= g_{21} = -\frac{\partial N}{\partial \lambda^1} = \kappa \frac{\Gamma(\eta + 2)}{\beta^{\eta+2}} \text{Li}_{\eta+1}(\xi), \\
g_{22} &= -\frac{\partial N}{\partial \lambda^2} = \kappa \frac{\Gamma(\eta + 1)}{\beta^{\eta+1}} \text{Li}_{\eta}(\xi) + \frac{\xi}{(1 - \xi)^2}.
\end{align*}
\]  

(3.27)

The only metric term that is influenced by the particles in the ground state is \(g_{22}\).

Then, the metric determinant takes the form

\[
g = \left(\frac{\kappa}{\beta^{\eta+2}}\right)^2 \bar{g}_b(\xi, \eta), \quad \text{where} \quad \bar{g}_b(\xi, \eta) \doteq \mathcal{A}(\xi, \eta) + \frac{\beta^{\eta+1}}{\kappa} \mathcal{A}_c(\xi, \eta).
\]  

(3.28)

Note that \(\bar{g}_b\) is dimensionless, \(\mathcal{A}\) is the same as in (3.20) and \(\mathcal{A}_c(x, \eta)\) is defined as

\[
\mathcal{A}_c(x, \eta) \doteq \det \begin{bmatrix} \Gamma(\eta + 3)\text{Li}_{\eta+2}(x) & 0 \\ \Gamma(\eta + 2)\text{Li}_{\eta+1}(x) & x/(1-x)^2 \end{bmatrix}.
\]  

(3.29)

If the \(\mathcal{A}_c\) term is ignored in (3.28) \(g\) reduces to \(g_0\) in (3.24). A graph for \(\bar{g}_b\) is presented in Fig. 3.5. We can see that the ground state term makes so that \(g\) grows faster than \(g_0\) with a rate of change that is increasing with \(\beta\). Also, in the graph it can be seen that for \(\eta = 1/2\), \(g_0\) diverges for \(\xi \to 1\) while it converges for \(\eta = 2\). In both cases, however, \(g\) diverges for positive values of \(\beta\).

The scalar curvature can be calculated from (2.124) so that for bosons we have

\[
R = -\frac{1}{2g^2} \left[ \frac{\kappa^3}{\beta^{3\eta+7}} B(\xi, \eta) + \frac{\kappa^2}{\beta^{2\eta+6}} B_c(\xi, \eta) \right] = -\frac{\beta^{\eta+1}}{2\kappa} \bar{R}_b(\xi, \eta),
\]  

(3.30)

where

\[
\bar{R}_b(\xi, \eta) \doteq \frac{B(\xi, \eta) + \frac{\beta^{\eta+1}}{\kappa} B_c(\xi, \eta)}{(\mathcal{A}(\xi, \eta) + \frac{\beta^{\eta+1}}{\kappa} \mathcal{A}_c(\xi, \eta))^2}.
\]  

(3.31)

\(B\) is the same as in (3.23) and \(B_c(x, \eta)\) is defined as.
Figure 3.4: Dimensionless quantity $\tilde{g}_b$, related to metric determinant as (3.28) for different values of $\beta$ (in units of $\kappa = 1$) for bosons in the values of $\eta = 1/2$ and $\eta = 2$. Figure previously published in [5].

Figure 3.5: Dimensionless quantity $\tilde{R}_b$, related to curvature as (3.30) for different values of $\beta$ (in units of $\kappa = 1$) for bosons in the values of $\eta = 1/2$ and $\eta = 2$. Figure previously published in [5].

$$B_c(x, \eta) \doteq \det \begin{bmatrix} \Gamma(\eta + 3)\text{Li}_{\eta+2}(x) & \Gamma(\eta + 2)\text{Li}_{\eta+1}(x) & \frac{x}{(1-x)^3} \\ \Gamma(\eta + 4)\text{Li}_{\eta+2}(x) & \Gamma(\eta + 3)\text{Li}_{\eta+1}(x) & 0 \\ \Gamma(\eta + 3)\text{Li}_{\eta+1}(x) & \Gamma(\eta + 2)\text{Li}_\eta(x) & \frac{x(x+1)}{(1-x)^3} \end{bmatrix}. \quad (3.32)$$

Once again, $R$ reduces to $R_0$ when the ground state term is ignored. A graph for the $\tilde{R}_b$ is presented in Fig. 3.5. From it we can see that $R_0$ diverges for $\xi \to 1$, however the addition of the ground state term eliminates the divergence, moreover the curvature converges to zero at positive values of $\beta$.

A representation of $R$ (3.30) in comparison to $R_0$ (3.24) in terms of $\xi$ and $\beta$ are
Figure 3.6: Contour plot of curvature, $R$ in (3.30), compared to their equivalents without the ground state terms, $R_0$ in (3.24). Represented as a function of $\xi$ and $\beta$ (units of $\kappa = 1$) in the relevant region for bosons and for the values of $\eta = \frac{1}{2}$ and $\eta = 2$. Figure previously published in [5].

Presented in Fig. 3.5. In accordance to the graphs of Fig. 3.5 we can see how the ground state term smooths the growth near $\xi = 1$, leads to a fast rise and falls to 0.

Having calculated the scalar curvature for both FD and BE we can compare them to the classical version. This will be done in the following section.

### 3.6 Classical limits

For a classical 3-dimensional ideal gas in the grand canonical ensemble the free-energy can be calculated as

$$F(\lambda) = \log \left[ \sum_{n=0}^{\infty} e^{-\lambda^2 n} \frac{V^n}{h^{3n} n!} \prod_{k=1}^{n} \int dp_k e^{-\lambda^2 \frac{p_k^2}{2m}} \right],$$

(3.33)
where \( m \) is the mass of particles in the gas. In the thermodynamical limit \( (N \to \infty) \) it can be written in the closed form

\[
F(\lambda) = -\kappa_{ig} \frac{\Gamma(\frac{3}{2})}{\beta^{3/2}} \xi , \quad \text{where} \quad \kappa_{ig} = \frac{V}{4\pi^2} \left( \frac{2m}{\hbar^2} \right)^{3/2},
\]

(3.34)

which is in accordance with the \( \kappa \) calculated for the quantum 3 dimensional particle in a box (see Table 1), here \( \xi \) is defined as in (3.15). For a comparison between (3.34) and its equivalent for quantum gases (3.17), it is useful to recall the series expansion for polylogarithms (3.13), around \( \xi = 0 \) meaning

\[
\text{Li}_\phi(\xi) = \xi + o(\xi).
\]

(3.35)

This implies that the free energy for both FD and BE gases (3.17) of a three dimensional particle in a box is equivalent – meaning it has the same dependence on the Lagrange multipliers – as the classical ideal gas in the limit \( \xi \to 0 \).

The free energy for the ideal gas leads to a FRIM calculated as

\[
g_{11} = -\frac{\partial^2 F}{\partial \lambda^1 \partial \lambda^1} = -\frac{\partial U}{\partial \lambda^1} = \kappa_{ig} \frac{\Gamma(7/2)}{\beta^{7/2}} \xi ,
\]

\[
g_{12} = g_{21} = -\frac{\partial^2 F}{\partial \lambda^2 \partial \lambda^1} = -\frac{\partial N}{\partial \lambda^1} = \kappa_{ig} \frac{\Gamma(5/2)}{\beta^{5/2}} \xi ,
\]

\[
g_{22} = \frac{\partial^2 F}{\partial \lambda^2 \partial \lambda^2} = -\frac{\partial N}{\partial \lambda^2} = \kappa_{ig} \frac{\Gamma(3/2)}{\beta^{3/2}} \xi.
\]

(3.36)

and this leads to the metric determinant and curvature

\[
g = \left( \frac{\kappa \xi}{\beta^{3/2}} \right)^2 \frac{\Gamma(3/2)\Gamma(5/2)}{\Gamma(3/2)} \text{ and } R = 0.
\]

(3.37)

Note that although the free energy of the ideal gas (3.34) reduces to the free energy of FD and BE statistics in the limit \( \xi \to 0 \), the curvature for FD (3.21) and BE (3.30) statistics do not converge to zero in the same limit. Instead, in this regime curvature for FD statistics is given by

\[
R_{FD} = \frac{\beta^{3/2}}{2\kappa} \lim_{x \to 0^+} \left[ \frac{B(x,1/2)}{(A(x,1/2))^2} \right] \approx -0.4987 \frac{\beta^{3/2}}{2\kappa}.
\]

(3.38)

The details for the calculation of the limit above are given in Appendix A.1. Equivalently, the curvature for BE statistics in the same limit, also to be fully calculated in
Appendix A.1 is

\[ R_{BE} = -\frac{\beta^{3/2}}{2\kappa} \lim_{x \to 0^+} \left[ \frac{B(x, 1/2) + \frac{\beta^{3/2}}{\kappa} B_c(x, 1/2)}{\left( A(x, 1/2) + \frac{\beta^{3/2}}{\kappa} A_c(x, 1/2) \right)^2} \right] \approx \left( \frac{0.6921 + 5.321 \frac{\beta^{3/2}}{\kappa}}{1.178 + 3.323 \frac{\beta^{3/2}}{\kappa}} \right)^2 \frac{\beta^{3/2}}{\kappa} . \]  

(3.39)

Therefore in order for the curvature in the FD and BE statistics to reduce to the one of the ideal gas it is not sufficient to take the low fugacity limit, rather one should take the high temperature limit \((\beta \to 0)\).

This result gives an interesting insight on the nature of quantum gases. The limit \(\xi \to 0\), or equivalently \(\lambda^2 = -\frac{\mu}{K_T} \to \infty\), refers to a sparse gas, as per constraints a high value of \(\lambda^2\) means a number of particles that is small for the scale of energy. In the work of Oshima and collaborators on the geometric description of quantum gases this result is interpreted as the curvature maintains quantum mechanics effects even in the “classical limit”. We argue against this interpretation by stating that a sparse quantum gas is not equivalent to a classical gas. Instead, one should expect the thermodynamics of a quantum gas to give the same results as in classical mechanics when the energy is bigger than the natural scale of energy of the system given by \(\kappa = 1\). Hence \(\beta \to 0\) is a more appropriate classical limit than only \(\xi \to 0\). Apart from this disagreement, it is important to say that this is a good example of how curvature gives information about the quantum structure of the system while free energy alone would not, stressing the importance and relevance of information geometry in thermodynamics.

It is also relevant to say that per (3.30) the vanishing curvature in the classical limit is fundamentally different from the convergence to zero in BE condensation. In the condensation regime the unitless curvature factor \(\bar{R}_b\) converges to zero as the limit of the fugacity \(\xi\) approaches one. In the classical limit, instead, \(\bar{R}_b\) is multiplied by the reciprocal of temperature \(\beta\) approaching zero.

\[16\] It is relevant to say here that, under appropriate units, the low fugacity limit for \(R_{FD}\) in (3.38) matches the result of [47]. On the other hand, the low fugacity limit for \(R_{BE}\) in (3.39) does not match [47], since the ground state contribution is still relevant for curvature even in that limit.

56
3.7 Conclusion

We obtained closed form expressions for the information metric of FD statistics in the continuous approximation (3.18), they also lead to the metric determinant and volume element (3.19) and scalar curvature (3.21). We also obtained similar expressions for the information metric of BE statistics in (3.27) leading to the metric determinant (3.28) and curvature (3.30). These results are more general than the ones previously obtained in the literature [46, 47, 48, 110] as they are calculated for a generic density of states exponent $\eta$, which opens for an IG description of several quantum mechanics models as illustrated in Table 3.1. Also it is the first time to the best of our knowledge that the ground state correction is being applied in these calculations. Therefore our analysis is carried out with the appropriate information metric for a system evolving into BE condensation.

As can be seen from Fig. 3.5 and Fig. 3.5, the qualitative behavior of curvature in the BE ideal gas is deeply altered by the addition of the ground state term. In the limit of fugacity approaching unity the scalar curvature converges to zero instead of diverging. This result is a counter-example to the hypothesis of curvature always diverging along a phase transition as BE condensation happens in that limit.

On that sense, we observe that (3.26) gives an accurate description of a system in which some particles are in the ground state and others are in the excited states, therefore the curvature does not diverge in a model that correctly describes both phases. In that understanding, our result suggests that the conjecture might be restated as: when the thermodynamical model only describes a single phase, a divergence in scalar curvature indicates where the probability model “breaks down”, which is usually identified as a phase transition.

The peculiarity of our proposed information geometric characterization of an ideal bosonic gas that condensates resides in the removal of the singular behavior of the scalar curvature of a manifold where regions with structurally different phases are no longer present. The absence of these distinct phases, in turn, is a consequence of having taken into account in an appropriate manner the ground state of the ideal bosonic gas. Despite its originality and broader applicability, our information geometric analysis is limited to non-interacting gases, further study is necessary on the physical meaning of the scalar curvature in information geometry. However, building on the work presented here, we are confident we will keep improving our comprehension of these fascinating quantum mechanical phenomena in future.
scientific efforts.
CHAPTER 4

Bose-Einstein statistics for a finite number of particles

4.1 Introduction

In most textbooks on statistical physics, e.g. [10, 113, 45], Bose-Einstein (BE) condensation is taught as one of the first examples of phase transitions. This phase transition is identified by the fact that there is a low value in temperature, namely the critical temperature, for which the specific heat of a non-interacting quantum gas of bosons is not analytical. Physically the condensation/phase transition happens because a thermodynamically relevant fraction of particles inhabits the quantum state of lowest energy, namely the ground state. The fraction of particles in the ground state itself also exhibits a non-analytical behaviour at the critical temperature, indicative of phase transition.

BE condensation was observed experimentally in 1995 [114, 115, 116]. In these experiments the number of particles ranged in the order of thousands [114] to millions [116]. This sprung interest in studies of BE statistics that do not rely on the thermodynamic limit (number of particles $N \to \infty$) [53, 54, 55, 56, 57, 58, 59, 60]. It is important to notice that the non-analytical behaviour mentioned previously is a consequence of the thermodynamic limit applied to BE statistics and therefore shouldn’t be expected for a finite number of particles.

In this chapter – which reviews the investigation done in [7] – I obtain more general results allowing for the studies of a broader range of quantum systems. Specifically, the present chapter studies quantum systems for which the density of states is proportional to a power of the energy\[^{17}\] $G(\epsilon) \propto \epsilon^\eta$. For a $D$-dimensional harmonically trapped gas $\eta = D - 1$, while the well-studied example of a gas in a box has $\eta = D/2 - 1$.

The major difficulty halting progress in this investigation comes from the lack of a closed form expression for the fugacity $\xi$ of a BE gas in terms of the number of particles $N$.

\[^{17}\]This relationship between the density of states and energy is known as dispersion relations. For further studies that investigate BE statistics for general dispersion relations see e.g. [117, 118, 5].
and temperature $T$, which is given by

$$N = \kappa \frac{\Gamma(\eta + 1)}{\beta \eta + 1} \text{Li}_{\eta + 1}(\xi) + \frac{\xi}{1 - \xi}, \quad (4.1)$$

where $\beta$ is the unit corrected inverse temperature, $\beta = \frac{1}{k_B T}$, where $k_B$ is the Boltzmann constant, and $\text{Li}$ refers to the polylogarithm family of functions — as defined in (3.13) — BE statistics assumes $\xi \in [0, 1)$. Note that for fixed $\beta$ and $\eta > 0$, $N$ is strictly increasing with $\xi$. Therefore $\xi(\beta, N)$ is well-defined as the inverse of (4.1). This, however, has not been written in closed analytical form to the best of my knowledge. This is fundamental for the study of BE statistics for a finite number of particles since the thermodynamical quantities that identify the phase transition — fraction of particles in the ground state and specific heat — can be written exactly in terms of $\beta$ and $\xi$, as we will see in the present chapter.

Most of the investigations done in BE statistics for finite number of particles is restricted to harmonically trapped gases, $\eta = D - 1$. Among the investigations that went beyond the study of harmonically trapped gases, it is important to mention the work of Jaouadi et. al. — that investigated the 3 dimensional gas of bosons in a power-law trap — and the work of Noronha — investigating the statistics for the Bose gas for a particular set of spaces and potentials, namely the particle in a box and a three-sphere. Both of these works, however, rely on approximations taken in the vicinity of critical temperature.

The present chapter differs from these aforementioned works since it does not assume a specific quantum system, rather here I presented an exact calculation of thermodynamical quantities in terms of $\beta$ and $\xi$ for arbitrary $\eta$ — later will be explained that one can only identify condensation for $\eta > 1$ — making it general. These calculations are presented for the first time here, to the best of my knowledge. Moreover, these results are complemented by the numerical calculation of $\xi(\beta, N)$, so all thermodynamical quantities can be calculated for a finite number of particles, allowing for reliable results irrespective of temperature.

A numerical implementation of $\xi(\beta, N)$ is found in my GitHub repository. It is important to point out that the calculation of fugacity cannot be simply implemented with the floating point arithmetic built in most computer languages, since for large $N$, $\xi$ is

---

18 Progress has been made by defining polyexponential functions as the inverse of (3.13). However, to the best of my knowledge, these polyexponentials were only written in closed analytical form for integer values of $\varphi$. 60
extremely close to unity. This difficulty is avoided by using the mpmath python library [111] which allows for arbitrary precision float point arithmetic — roughly speaking, all quantities can be calculated with an arbitrary number of decimal places.\footnote{The graphs presented here are based on a calculation with 50 decimal places.}

From the fugacity obtained numerically, one can calculate the fraction of particles in the ground state, specific heat and specific heat first derivative and compare to the results obtained for them in the thermodynamic limit, available in [10, 113, 45, 117]. For transparent presentation of the results, graphs for all the calculated quantities are presented for number of particles ranging from $N = 10^2$ to $N = 10^7$, and its comparison to the thermodynamic limit.

The layout of the present chapter is as follows: Sec. 4.2 will review the statistical mechanical description of quantum gases in terms of $\beta$ and $\xi$ and obtain the parameter $\eta$ for some well known models. Sec. 4.3 will present the description of Bose gases in the thermodynamic limit for different values of $\eta$ and show how these values affect the convergence of critical temperature and the existence of phase transition identified by non analytical behaviour. Sec. 4.4 presents exact results for the fraction of particles in the ground state and specific heat in terms of $\beta$ and $\xi$. These quantities are plotted in terms of $\beta$ and $N$ from the numerical implementation of $\xi(\beta, N)$ found in [61]. These calculations do not specify a value of $\eta$ and the graphs are presented for $\eta = 1/2$ and $\eta = 2$.

### 4.2 Spectrum and density of states

Quantum statistical mechanics consists of assigning probability distributions for the space of occupancy number of each quantum state — also referred to as Fock space. Mathematically this means assigning a probability $\rho(x)$ where $x = \{x_i\}$ in which $i$ enumerates the quantum states. In this description, $x_i$ is the number of particles occupying the state $i$ — for bosons $x_i$ takes positive integer values, $x_i \in \{0, 1, 2, \ldots\}$ — and each quantum state is associated to an energy value $\epsilon_i$. The relationship between $i$ and $\epsilon_i$ is also referred to as the spectrum and is obtained through regular methods in quantum mechanics [120], that means $\epsilon_i$ are the eigenvalues of the Hamiltonian operator $\hat{H}$ defined as

$$\hat{H} \equiv \frac{\hat{p}^2}{2m} + \phi(q),$$

(4.2)
where \( q = \{q_1, q_2, \ldots q_D\} \) are the system’s position coordinates — which, for simplicity, are assumed to be rectangular — \( \hat{p} \) is the momentum identified as \( \hat{p} = -i\hbar \sum_{\mu=1}^{D} e_\mu \frac{\partial}{\partial q_\mu} \) — where \( e_\mu \) refers to the unit vector in the direction of the \( \mu \)-th dimension — and \( \phi(q) \) is the potential, \( m \) refers to the mass of the confined particles, and \( \hbar \) is the reduced Planck constant. For mathematical simplicity, one can shift the energy spectrum so that the energy of the ground state is zero, \( \min_i \epsilon_i = 0 \).

In the grand canonical ensemble \( \rho(x) \) is assigned as the maximum entropy distribution constrained on the total number of particles, \( N \doteq \sum_i \langle x_i \rangle \), and the system’s internal energy, \( U \doteq \sum_i \epsilon_i \langle x_i \rangle \). This leads to the grand canonical Gibbs distribution

\[
\rho(x|\beta, \xi) \propto \prod_i e^{-\beta \epsilon_i x_i} \xi^{x_i}, \tag{4.3}
\]

where, as mentioned, \( \beta = \frac{1}{k_B T} \) and the fugacity \( \xi \) is related to the chemical potential \( \mu \) by \( \xi = e^{\beta \mu} \). The number of particles \( N \) and the internal energy \( U \) can be calculated from (4.3) as

\[
N = \sum_i \frac{\xi}{e^{\beta \epsilon_i} - \xi}, \tag{4.4a}
\]

\[
U = \sum_i \frac{\epsilon_i \xi}{e^{\beta \epsilon_i} - \xi}. \tag{4.4b}
\]

The details of how the Gibbs distribution (4.3) is derived as well as how it leads to (4.4) was presented in section 3.2.

In order to calculate \( N \) and \( U \) in terms of \( \beta \) and \( \xi \) one needs the full spectrum of energies. I am not aware of any calculations of the series (4.4) in closed form for a general spectrum nor any calculation directly from (4.4) for physically relevant systems.

Because of this, generally the study of quantum statistics mechanics relies on treating the spectrum as a continuum. This justified when the energy of the system is much larger than the differences of energy in the spectrum, \( U \gg \max_{i,j} |\epsilon_i - \epsilon_j| \). In such approximation the summations in (4.4) can be substituted as

\[
\sum_i \rightarrow \int d\epsilon \ G(\epsilon) = \kappa \int d\epsilon \ e^{\beta \epsilon}, \tag{4.5}
\]

62
where $\kappa$ and $\eta$ are parameters to be calculated from the full spectrum of energies, as it will be presented below. Note that $\kappa$ has units of $[\text{energy}]^{(\eta+1)}$ — or $\kappa^{1/(\eta+1)}$ has units of energy — one can choose a system of units in which $\kappa = 1$, or, equivalently, use $\kappa^{1/(\eta+1)}$ as the unit of energy given by the system.

As a first example of the continuous approximation, one can study the $D$-dimensional gas trapped in a box of edge length $L$. This is related to a potential $\phi$ of the form: $\phi(q) = 0$ if $\max_{\mu} |q_{\mu}| \leq \frac{L}{2}$ and $+\infty$ otherwise. In this potential, the quantum states will be enumerated by $D$ integers — $i = \{n_1, n_2, \ldots, n_D\}$ — and the spectrum of energies is given by

$$\epsilon_i = \frac{\pi^2 \hbar^2}{2mL^2} \sum_{\mu=0}^{D} n_{\mu}^2.$$  (4.6)

In this case the continuous approximation yields

$$\eta = \frac{D}{2} - 1 \quad \text{and} \quad \kappa = \frac{g_s V}{\Gamma(D/2)} \left( \frac{m}{2\pi \hbar^2} \right)^{D/2},$$  (4.7)

where $V$ is the volume of the box, $V = L^D$, and $g_s$ is the multiplicity of energy levels of a particle with spin $s$, $g_s = 2s + 1$.

Another useful example is the $D$-dimensional gas in a harmonic trap, meaning a potential of the form: $\phi(q) = \frac{m}{2} \omega^2 \sum_{\mu} q_{\mu}^2$. In this potential, the quantum states will be enumerated by $D$ positive integers — $i = \{n_1, n_2, \ldots, n_D\}$ — and the spectrum of energies is given by

$$\epsilon_i = \hbar \omega \sum_{\mu=0}^{D} n_{\mu},$$  (4.8)

the extra term $\frac{1}{2} \hbar \omega D$ in the energy of a harmonic trapped particle is ignored, in order to assign zero energy to the ground state — $n_{\mu} = 0$ for all $\mu$. In this case, the continuous approximation yields

$$\eta = D - 1 \quad \text{and} \quad \kappa = \frac{g_s}{\Gamma(D)} \left( \frac{1}{\hbar \omega} \right)^D.$$  (4.9)

As final example — useful to describe a more general set of quantum systems — one can calculate the values of $\kappa$ and $\eta$ for a potential of the form $\phi(q) = \phi_0 \sum_{\mu} |2q_{\mu}|^r$. The
density of states can be calculated in a semi-classical manner from the phase space volume, meaning

\[
\int_0^E \frac{dG(\epsilon)}{d\epsilon} = \frac{1}{(2\pi\hbar)^D} \int_{A(E)} d^Dp \, d^Dq ,
\]  
(4.10)

where \( A(E) = \{ (p, q) : \frac{1}{2m} \sum_\mu p_\mu^2 + \phi(q) \leq E \} \) as presented in [118, 122, 59]. This yields

\[
\eta = \frac{D}{2} + \frac{D}{\tau} - 1 \quad \text{and} \quad \kappa = \frac{g_s V}{\Gamma(D/2 + D/\tau)} \left( \frac{2m}{\pi \hbar^2} \right)^{D/2} \left[ \frac{\Gamma(1/\tau + 1)}{\sqrt{2\phi_0}} \right]^D .
\]  
(4.11)

It is straightforward to see that when the potential \( \phi(x) \) is a harmonic oscillator — \( \tau = 2 \) and \( (2L^2)^2 \phi_0 = \frac{m}{\tau} \omega^2 \) — the values of \( \eta \) and \( \kappa \) in (4.11) match those of (4.9). On the same note, when \( \phi(x) \) reduces to the potential of a particle in a box — \( \tau \to \infty \) and \( \phi_0 = 1 \) — the values of \( \eta \) and \( \kappa \) in (4.11) match those of (4.7).

Having the values of \( \eta \) and \( \kappa \) from the energy spectrum, one can calculate the number of particles and the internal energy (4.4) using the continuous approximation (4.5) thus obtaining

\[
N = \kappa \frac{\Gamma(\eta + 1)}{\beta^{\eta+1}} \text{Li}_{\eta+1}(\xi) + n_0 , \quad \text{where} \quad n_0 = \frac{\xi}{1 - \xi} ,
\]  
(4.12a)

\[
U = \kappa \frac{\Gamma(\eta + 2)}{\beta^{\eta+2}} \text{Li}_{\eta+2}(\xi) .
\]  
(4.12b)

Note that \( n_0 \) is the number of particles in the ground state, which needs to be added ad hoc since the continuous approximation (4.5) assigns no particle in the ground state, \( G(0) = 0 \).

It can also be seen that \( n_0 \) is the term in \( N \) equivalent to the zero energy state in (4.4a).

As will be discussed in the remainder of the present chapter, the addition of this term is fundamental for the study of BE condensation.

Before studying the thermodynamical consequences of (4.12), two important properties of polylogarithms need attention for future use. First, it is useful to note that from (3.13) one obtains

\[
\frac{\partial}{\partial y} \text{Li}_\varphi(y) = \frac{1}{y} \text{Li}_{\varphi-1}(y) .
\]  
(4.13)

Second, as deduced by Cohen et. al. [123], for non-integer \( \varphi \) polylogarithms can be written

\[\text{[Footnote 20]}\]
as a series
\[
\text{Li}_\varphi(y) = \Gamma(1 - \varphi)(-\log(y))^{\varphi -1} + \sum_{k=0}^{\infty} \zeta(\varphi - k) \frac{(\log y)^k}{k!},
\]
(4.14)
where \(\zeta\) refers to the Riemann’s zeta function, \(\zeta(\varphi) = \sum_{k=1}^{\infty} k^{-\varphi}\). This series expression is valid for \(|\log y| < 1\). A similar expression of integer \(\varphi\) is also found in [123]. Note that for \(\varphi > 1\) it implies that \(\lim_{y \to 1^-} \text{Li}_\varphi(y) = \zeta(\varphi)\) while \(\varphi \leq 1\) implies \(\lim_{y \to 1^-} \text{Li}_\varphi(y) = \infty\).

### 4.3 BE statistics in the thermodynamic limit

This section will describe the phase transition for BE statistics in the thermodynamic limit by presenting the non-analytical form of the fraction of particles in the ground state and the specific heat. Appendix B.1 will show how these quantities derive from the thermodynamical quantities presented in this section follow from the calculations made for finite \(N\) in Sec. 4.4. These calculations will leave \(\eta\) undetermined, they reduce to those found in textbooks [10, 113, 45] when \(\eta = \frac{1}{2}\).

When studying BE condensation, \(\beta_c\) defined as
\[
\beta_c \doteq \left[ \frac{\Gamma(\eta + 1)}{\kappa \frac{N}{1}} \lim_{\xi \to 1^-} \text{Li}_{\eta+1}(\xi) \right]^{\frac{1}{\eta +1}},
\]
(4.15)
is identified as the inverse critical temperature\(^{21}\). This definition is motivated as the temperature for which \(\xi\) goes to 1 when the ground state particles are ignored in (4.12a). From (4.14) it follows that for \(\eta \leq 0\) \(\beta_c\) diverges — or the critical temperature goes to the absolute zero. When \(\eta\) is positive \(\beta_c\) converges and it yields \(\beta_c = \left[ \kappa \frac{\Gamma(\eta + 1)}{N} \zeta(\eta + 1) \right]^{\frac{1}{\eta +1}}\). For the discussion presented here, one can assume \(\eta > 0\), guaranteeing a positive critical temperature. Note that this means, per Sec. 4.2, a 1 or 2 dimensional Bose gas in a box will not have a positive critical temperature. Similarly, a 1 dimensional Bose gas in a box has divergent \(\beta_c\).

Two assumptions are applied when studying the Bose gas in the thermodynamic limit. These can be summarized as

- **For \(\beta < \beta_c\)**: treat the calculations of thermodynamical quantities as if \(n_0 = 0\),

---

\(^{21}\)Other definitions for critical temperature were studied before, see e.g. [53, 124, 125]. These studies however are beyond the scope of the present tome.
• For $\beta \geq \beta_c$ : treat the calculations of thermodynamical quantities as if $\xi = 1$.

This leads to a non-analytical behavior in thermodynamical quantities as will be presented below. Before presenting the results in the thermodynamic limit, they will be calculated for a finite number of particles in Sec. 4.4 and the thermodynamic limit will be taken in Appendix B.1. It will also present comparisons in the form of graphs between the quantities calculated in thermodynamic limit and for a finite number of particles.

The fraction of particle in the ground state is given by

$$\tilde{n}_0 \frac{N}{\bar{N}} = \begin{cases} 
0 & \text{for } \beta < \beta_c \\
1 - \left(\frac{\beta}{\beta_c}\right)^{-(\eta+1)} & \text{for } \beta \geq \beta_c ,
\end{cases} \quad (4.16)$$

for the remainder of the present chapter I will use the tilde notation as above to mean that the quantity ($\tilde{n}_0$, $\tilde{c}_v$, and $\tilde{\xi}$) is calculated in the thermodynamic limit.

Similarly the specific heat\footnote{In many texts $c_v$ is referred to as specific heat at constant volume, as explained in Sec. 4.2, $\kappa$ absorbs the dependence with volume. That is why our derivatives in (4.17) and (4.20) are taken under constant $\kappa$.} defined as $c_v \equiv \frac{1}{N} \left(\frac{\partial U}{\partial T}\right)_{N,\kappa} = -k_B \frac{\beta^2}{N} \left(\frac{\partial U}{\partial \beta}\right)_{N,\kappa}$ yields, in the thermodynamic limit

$$\tilde{c}_v \frac{k_B}{\kappa} = \begin{cases} 
(\eta + 2)(\eta + 1) \frac{\text{Li}_{\eta+2}(\tilde{\xi})}{\text{Li}_{\eta+1}(\tilde{\xi})} - (\eta + 1)^2 \frac{\text{Li}_{\eta+1}(\tilde{\xi})}{\text{Li}_{\eta}(\tilde{\xi})} & \text{for } \beta < \beta_c \\
(\eta + 2)(\eta + 1) \frac{\zeta(\eta + 2)}{\zeta(\eta + 1)} \left(\frac{\beta}{\beta_c}\right)^{-(\eta+1)} & \text{for } \beta \geq \beta_c
\end{cases} , \quad (4.17)$$

Where $\tilde{\xi}$ is the fugacity in the thermodynamic limit for $\beta < \beta_c$, which can be obtained as the solution to (4.12a) with the ground state particles ignored and substituting $\beta_c$ defined in (4.15). Namely $\tilde{\xi}(\beta)$ is the solution to

$$\text{Li}_{\eta+1}(\tilde{\xi}) = \zeta(\eta + 1) \left(\frac{\beta}{\beta_c}\right)^{\eta+1} . \quad (4.18)$$

Note that the limit $\beta \to 0$ leads to $\tilde{\xi} \to 0$ and for $\beta \to \beta_c^-$ it follows that $\tilde{\xi} \to 1$. In order to study the non-analytical behavior of $\tilde{c}_v$, we define the quantity

$$\frac{\Delta \tilde{c}_v}{k_B} \equiv \lim_{\beta \to \beta_c^-} \tilde{c}_v - \lim_{\beta \to \beta_c^+} \tilde{c}_v = (\eta + 1)^2 \lim_{\xi \to 1^-} \frac{\text{Li}_{\eta+1}(\xi)}{\text{Li}_{\eta}(\xi)} , \quad (4.19)$$

$$\Delta \tilde{c}_v \frac{k_B}{\kappa} \equiv \lim_{\beta \to \beta_c^-} \tilde{c}_v - \lim_{\beta \to \beta_c^+} \tilde{c}_v = (\eta + 1)^2 \lim_{\xi \to 1^-} \frac{\text{Li}_{\eta+1}(\xi)}{\text{Li}_{\eta}(\xi)} , \quad (4.19)$$
which is the discontinuity gap of \( \tilde{c}_v \) at \( \beta = \beta_c \). For \( \eta \leq 1 \), it follows that \( \lim_{\xi \to 1^-} \text{Li}_\eta(\xi) = \infty \) and \( \lim_{\xi \to 1^-} \text{Li}_{\eta+1}(\xi) = \zeta(\eta + 1) \) therefore \( \Delta \tilde{c}_v = 0 \). While for \( \eta > 1 \), \( \tilde{c}_v \) is discontinuous in \( \beta_c \), yielding \( \Delta \tilde{c}_v = (\eta + 1)^2 \zeta(\eta + 1) / \zeta(\eta) \).

One can further study the derivative of \( c_v \), using the unitless quantity \( \frac{1}{k_B \beta} \left( \frac{\partial c_v}{\partial T} \right)_{N,\kappa} = -\frac{\beta}{k_B} \left( \frac{\partial c_v}{\partial \beta} \right) \) obtained from differentiating \( \text{(4.17)} \), yielding

\[
\frac{1}{k_B^2 \beta} \frac{\partial \tilde{c}_v}{\partial T}_{N,\kappa} = \begin{cases} 
(\eta + 2)(\eta + 1)^2 \text{Li}_{\eta+2}(\tilde{\xi}) / \text{Li}_{\eta+1}(\tilde{\xi}) - (\eta + 1)^2 \text{Li}_{\eta+1}(\tilde{\xi}) / \text{Li}_{\eta}(\tilde{\xi}) & \text{for } \beta < \beta_c \\
-(\eta + 1)^3 \frac{(\text{Li}_{\eta+1}(\tilde{\xi}))^2 \text{Li}_{\eta-1}(\tilde{\xi})}{(\text{Li}_\eta(\tilde{\xi}))^3} & \text{for } \beta \geq \beta_c 
\end{cases}.
\]

(4.20)

Similarly to \( \text{(4.19)} \), one can define the discontinuity gap of the derivative of \( \tilde{c}_v \) at \( \beta = \beta_c \)

\[
\frac{1}{k_B^2 \beta} \Delta \frac{\partial \tilde{c}_v}{\partial T}_{N,\kappa} \doteq \lim_{\beta \to \beta_c^+} \frac{1}{k_B^2 \beta} \frac{\partial \tilde{c}_v}{\partial T}_{N,\kappa} - \lim_{\beta \to \beta_c^-} \frac{1}{k_B^2 \beta} \frac{\partial \tilde{c}_v}{\partial T}_{N,\kappa}.
\]

(4.21)

Calculating the quantity \( \Delta \left( \frac{\partial c_v}{\partial T} \right)_{N,\kappa} \) for \( \eta \leq 0 \) is not necessary, since \( \beta_c \) diverges, on the same understanding calculating \( \Delta \left( \frac{\partial c_v}{\partial T} \right)_{N,\kappa} \) is not representative for \( \eta > 1 \) since \( \tilde{c}_v \) is already discontinuous. This calculation will focus on values \( 0 < \eta \leq 1 \). In that regime, it follows from \( \text{(4.20)} \) that the limit from the right is given by

\[
\lim_{\beta \to \beta_c^+} \frac{1}{k_B^2 \beta} \frac{\partial \tilde{c}_v}{\partial T}_{N,\kappa} = (\eta + 2)(\eta + 1)^2 \zeta(\eta + 2) / \zeta(\eta + 1) .
\]

(4.22)

In order to calculate the equivalent limit from the left, one has to recall the series expansion in \( \text{(4.14)} \). Note that, for the values of \( \eta \) of interest, \( \text{Li}_\eta(\xi) \) scales as \( (\log(\xi))^{\eta - 1} \) as \( \xi \to 1^- \), while \( \text{Li}_{\eta-1}(\xi) \) scales as \( (\log(\xi))^{\eta - 2} \) while \( \text{Li}_{\eta+1}(\xi) \) and \( \text{Li}_{\eta+2}(\xi) \) converge to \( \zeta(\eta + 1) \) and \( \zeta(\eta + 2) \), respectively. Substituting the series expansion in \( \text{(4.20)} \) it follows that

\[
\lim_{\beta \to \beta_c^-} \frac{1}{k_B^2 \beta} \frac{\partial \tilde{c}_v}{\partial T}_{N,\kappa} = (\eta + 2)(\eta + 1)^2 \zeta(\eta + 2) / \zeta(\eta + 1) - (\eta + 1)^3(\zeta(\eta + 1))^2 / \zeta(1-\eta) \Gamma(2-\eta) \lim_{\xi \to 1^-} (-\log \xi)^{1-2\eta}.
\]

(4.23)
For $0 < \eta < 1/2$ the exponent in the last factor in (4.23) is positive, therefore $\Delta \left( \frac{\partial \tilde{c}_v}{\partial T} \right)_{N, \kappa} = 0$. For $1/2 < \eta \leq 1$ the exponent in the last factor in (4.23) is negative, therefore $\Delta \left( \frac{\partial \tilde{c}_v}{\partial T} \right)_{N, \kappa} = \infty$.

In the particular case of $\eta = 1/2$ — as a 3-dimensional gas trapped in a box — the exponent vanishes, therefore

$$
\lim_{\beta \to \beta_c} \frac{1}{k_B^2 \beta} \left( \frac{\partial \tilde{c}_v}{\partial T} \right)_{N, \kappa} = \left( \frac{45}{8} \right) \frac{\Gamma(5/2)}{\Gamma(3/2)} - \left( \frac{3}{2} \right)^3 \frac{\Gamma(3/2)}{\Gamma(1/2)}^2 \approx -0.77726 .
$$

(4.24)

Leading to a convergent $\Delta \left( \frac{\partial \tilde{c}_v}{\partial T} \right)_{N, \kappa} \approx 3.6657 k_B^2 \beta_c$ in accordance to [10, 113].

A summary for the convergence of $\beta_c$ along with the discontinuities of $\tilde{c}_v$ and $\left( \frac{\partial \tilde{c}_v}{\partial T} \right)_{N, \kappa}$, in terms of $\eta$ are presented in Table 4.1. With the study of the thermodynamic limit for general values of $\eta$ done, the following section studies BE statistics for a finite number of particles.

### 4.4 BE statistics for a finite number of particles

In order to appropriately study BE condensation in terms of $N$, one needs to express the thermodynamical quantities of interest — namely the fraction of particles in the ground state, specific heat and its derivative — in a manner that is appropriate to compare to the critical temperature. Since $\beta_c$ in (4.15) is defined in terms of $N$, one has to write $n_0$ and $U$ in terms of $\beta$ and $N$, since these were written in (4.12) in terms of $\beta$ and $\xi$ it would suffice to obtain $\xi$ as a function of $\beta$ and $N$. From (4.12a) it is straightforward to see that $N$ is strictly increasing with $\xi \in [0, 1)$ therefore $\xi(\beta, N)$ is well defined as the inverse of (4.12a).

To the best of my knowledge, $\xi(\beta, N)$ has never been written in closed analytical form. However as the inverse of a strictly increasing function $\xi(\beta, N)$ can be implemented through simple numerical algorithms. An implementation of it can be seen in the IGQG python library, available in my GitHub repository [61]. This implementation is based on the mpmath library [111] that allows for calculations of arbitrary precision. All thermodynamical quantities of interest can be exactly written in terms of $\xi(\beta, N)$ — which will be written only as $\xi$ in this section for simpler notation — Graphs for the thermodynamical quantities obtained from this implementation will be presented below. It is observed that for finite $N$ the non analytical behaviour disappears, which is in accordance with the fact that $U$ and $N$

\footnote{To the best of my knowledge These calculations were first published in [7]}

68
In this table we show the relationship between the convergence of polylogarithms (in terms of the density of states exponent $\eta$) to the convergence of the critical temperature, $\beta_c$ in (4.15). It is also presented how, as a consequence, the value of $\eta$ affects the continuity of specific heat, $\tilde{c}_v$ in (4.17), and of its derivative in (4.21). Table previously published in [7]. Table previously published in [7].

<table>
<thead>
<tr>
<th>$\lim_{\xi \to 1^-} \text{Li}_{\eta+1}(\xi)$</th>
<th>$\eta \leq 0$</th>
<th>$0 &lt; \eta &lt; \frac{1}{2}$</th>
<th>$\eta = \frac{1}{2}$</th>
<th>$\frac{1}{2} &lt; \eta \leq 1$</th>
<th>$\eta &gt; 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lim_{\xi \to 1^-} \text{Li}_\eta(\xi)$</td>
<td>Divergent (+\infty)</td>
<td>Divergent (+\infty)</td>
<td>Convergent $\zeta(\eta + 1)$</td>
<td>Convergent $\zeta(\eta)$</td>
<td>$\zeta(\eta)$</td>
</tr>
<tr>
<td>$\beta_c$</td>
<td>Divergent (+\infty)</td>
<td>Convergent — see (4.15)</td>
<td>Continuous at $\beta_c$</td>
<td>Discontinuous at $\beta_c$</td>
<td>$\beta_c$</td>
</tr>
<tr>
<td>$\tilde{c}_v$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
</tr>
<tr>
<td>$\Delta \tilde{c}_v$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
</tr>
<tr>
<td>$\left( \frac{\partial \tilde{c}<em>v}{\partial \beta} \right)</em>{N,\kappa}$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
</tr>
<tr>
<td>$\Delta \left( \frac{\partial \tilde{c}<em>v}{\partial \beta} \right)</em>{N,\kappa}$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
<td>Continuous at $\beta_c$</td>
</tr>
<tr>
<td>$\approx 3.66 \frac{k_B^2}{\beta_c}$</td>
<td>$+\infty$</td>
<td>$+\infty$</td>
<td>$+\infty$</td>
<td>$+\infty$</td>
<td>$+\infty$</td>
</tr>
</tbody>
</table>

Table 4.1: In this table we show the relationship between the convergence of polylogarithms (in terms of the density of states exponent $\eta$) to the convergence of the critical temperature, $\beta_c$ in (4.15). It is also presented how, as a consequence, the value of $\eta$ affects the continuity of specific heat, $\tilde{c}_v$ in (4.17), and of its derivative in (4.21). Table previously published in [7]. Table previously published in [7].
in (4.12) are continuous for positive \( \beta \) and \( 0 \leq \xi < 1 \).

The fraction of particles in the ground state can be obtained by dividing (4.12a) by the number of particles \( N \) and substituting \( \beta_c \) as in (4.15) obtaining

\[
\frac{n_0}{N} = 1 - \frac{\text{Li}_{\eta+1}(\xi)}{\zeta(\eta + 1)} \left( \frac{\beta}{\beta_c} \right)^{- (\eta + 1)}. \tag{4.25}
\]

Above it is supposed a value of \( \eta \) for which \( \beta_c \) converges, hence \( \lim_{\xi \to 1^-} \frac{\text{Li}_{\eta+1}(\xi)}{\zeta(\eta + 1)} \) is taken to be \( \zeta(\eta + 1) \). Graphs for \( \frac{n_0}{N} \) are presented for the 3-dimensional gas in a box (\( \eta = \frac{1}{2} \)) and the 3-dimensional harmonically trapped gas (\( \eta = 2 \)) in Fig. 4.1 with a comparison to the fraction of particles calculated in the thermodynamic limit (4.17).

The specific heat, \( c_v \), can be calculated from the direct differentiation of \( U \) in (4.12b), obtaining

\[
c_v = \frac{1}{k_B} \left( \frac{\partial U}{\partial T} \right)_{N,\kappa} = -k_B \frac{\beta^2}{N} \left( \frac{\partial U}{\partial \beta} \right)_{N,\kappa},
\]

where \( \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \) can be obtained from the implicit differentiation of \( N \) in (4.12a) with respect to \( \beta \), yielding

\[
\frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} = \frac{\kappa \Gamma(\eta + 2)}{\beta^{\eta+1}} \text{Li}_{\eta+1}(\xi) + \frac{\xi}{(1 - \xi)^2}. \tag{4.27}
\]

Note that the ground state contribution to \( c_v \) appears in the second term of the denominator of (4.27). Graphs for \( c_v \) are presented for \( \eta = \frac{1}{2} \) and \( \eta = 2 \) in Fig. 4.2. The comparison to the thermodynamic limit is obtained from (4.17). For \( \beta \geq \beta_c \) it follows that \( \tilde{c}_v \) can be calculated directly in terms of \( \beta \) directly, for \( \beta < \beta_c \) the plotted values of \( \tilde{c}_v \) are based on the numerical implementation of the solution of (4.18) also found in [61]. It is particularly interesting to notice, in Fig. 4.2, how the discontinuity in \( \tilde{c}_v \) for \( \eta = 2 \) is approached from the continuous values calculated for finite \( N \).

To further compare the study of BE statistics for finite \( N \) to the one in the thermodynamic limit, it is important to study the derivative of \( c_v \). Again, this can be done by the study of the unitless quantity \( \frac{1}{k_B^2} \left( \frac{\partial c_v}{\partial T} \right)_{N,\kappa} = -\frac{\beta}{k_B} \left( \frac{\partial c_v}{\partial \beta} \right)_{N,\kappa} \) which can be obtained from
Figure 4.1: Fraction of particles in the ground state for $\eta = 1/2$ and $\eta = 2$. On both cases it can be seen how the fraction of particle in the ground state, for finite $N$, approach, smoothly, the non-analytical curve for the thermodynamic limit. Figure previously published in [7].
Figure 4.2: Specific heat of a Bose gas (4.26) for $\eta = \frac{1}{2}$ and $\eta = 2$. In the first picture it is seen that the quantity is continuous, albeit not smooth, as $\beta$ approaches $\beta_c$ in the thermodynamic limit. In the second one it is seen how in the thermodynamic limit, the value diverges is discontinuous when $\beta$ approaches $\beta_c$ from the left. On both cases it can be seen how the non-analytical curve is approached from the curves of specific heat for finite $N$. Figure previously published in [7].
differentiating (4.26) yielding

\[
\frac{1}{k_B^2 \beta} \left( \frac{\partial \tilde{c}_v}{\partial T} \right)_{N,\kappa} = \frac{\kappa}{N} \frac{\Gamma(\eta + 2)}{\beta^{\eta+1}} \left[ (\eta + 2)(\eta + 1) \text{Li}_{\eta+2}(\xi) - \frac{\beta}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \right] 2(\eta + 1) \text{Li}_{\eta+1}(\xi)
\]

\[
+ \left( \frac{\beta}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \right)^2 (\text{Li}_\eta(\xi) - \text{Li}_{\eta+1}(\xi)) + \frac{\beta^2}{\xi} \left( \frac{\partial^2 \xi}{\partial \beta^2} \right)_{N,\kappa} \text{Li}_{\eta+1}(\xi) \right],
\]

(4.28)

where \( \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \) was already calculated in (4.27) and \( \left( \frac{\partial^2 \xi}{\partial \beta^2} \right)_{N,\kappa} \) can similarly be obtained from the second implicit differentiation of \( N \) in (4.12a) with respect to \( \beta \), yielding

\[
\frac{1}{\xi} \left( \frac{\partial^2 \xi}{\partial \beta^2} \right)_{N,\kappa} = - \left[ \kappa \frac{\Gamma(\eta + 3)}{\beta^{\eta+3}} \text{Li}_{\eta+1}(\xi) - \frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} 2k \frac{\Gamma(\eta + 2)}{\beta^{\eta+2}} \text{Li}_\eta(\xi)
\]

\[
+ \left( \frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \right)^2 \kappa \frac{\Gamma(\eta + 1)}{\beta^{\eta+1}} (\text{Li}_{\eta-1}(\xi) - \text{Li}_\eta(\xi))
\]

\[
+ 2 \left( \frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \right)^2 \frac{\xi^2}{(1 - \xi)^3} \right] \times \left[ \kappa \frac{\Gamma(\eta + 1)}{\beta^{\eta+1}} \text{Li}_\eta(\xi) + \frac{\xi}{(1 - \xi)^2} \right]^{-1}.
\]

(4.29)

Graphs for the unitless quantity \( \frac{1}{k_B^2 \beta} \left( \frac{\partial \xi}{\partial T} \right)_{N,\kappa} \) are presented for \( \eta = 1/2 \) and \( \eta = 2 \) in Fig. 4.3.

An interesting non intuitive behaviour becomes clear in Fig. 4.3. For \( \eta = 1/2 \), it can be seen that the value of \( \frac{1}{k_B^2 \beta} \left( \frac{\partial \xi}{\partial T} \right)_{N,\kappa} \) grows smaller than the minimum possible value obtained in the thermodynamic limit, given as \( \lim_{\beta \to \beta_c} \frac{1}{k_B^2 \beta} \left( \frac{\partial \xi}{\partial T} \right)_{N,\kappa} \) in (4.24). Such behavior is not observed for \( \eta < 1/2 \) — since at those values \( \left( \frac{\partial \xi}{\partial T} \right)_{N,\kappa} \) is continuous — it is also not observed for \( \eta > 1/2 \) — since, per (4.23), it follows that \( \lim_{\beta \to \beta_c} \frac{1}{k_B^2 \beta} \left( \frac{\partial \xi}{\partial T} \right)_{N,\kappa} \) goes to negative infinite.

This implies that the thermodynamic limit, as presented in Sec. 4.3, misses interesting physical behaviour. Mainly, the calculation of \( \tilde{c}_v \) was made assuming that there are no particles in the ground state for \( \beta < \beta_c \) – as is also the case for all calculations made in Sec. 4.3. The fact that \( \left( \frac{\partial \xi}{\partial T} \right)_{N,\kappa} \) obtained this way is strictly decreasing for \( \beta < \beta_c \) indicated that the discontinuity at \( \beta_c \) is approached from above. The behaviour found in Fig. 4.3 indicates that accounting for \( n_0 \) above the critical temperature leads to a smaller value of \( \left( \frac{\partial \xi}{\partial T} \right)_{N,\kappa} \).
Figure 4.3: Graphs for $\left(\frac{\partial c_v}{\partial T}\right)_{N,\kappa}$ of a Bose gas for $\eta = 1/2$ and $\eta = 2$. On the first image it is seen that the quantity becomes smaller than the one expected from calculations in the thermodynamic limit. On the second one it is seen how in the thermodynamic limit, the value diverges to $-\infty$ when $\beta$ approaches $\beta_c$ from the left. Figure previously published in [7].
Therefore the discontinuity found in the thermodynamic limit is approached from below, not above.

Interestingly, this result is supported by analytical calculations. For \( \eta = \frac{1}{2} \), the minimal value of \( \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} \) — calculated without assuming \( n_0 = 0 \) below \( \beta_c \) — is related to \( N \) as

\[
z(N) \doteq \min_{\beta} \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} = z_m + \bar{z} N^{-1/3} + o(N^{-1/3}) ,
\]

where

\[
z_m \approx -0.97337 \quad \text{and} \quad \bar{z} \approx 3.5881 ,
\]

and where \( o \) stands for the smaller order notation, \( \lim_{N \to \infty} \frac{o(N^{-1/3})}{N^{-1/3}} = 0 \). Note that, as expected from Fig. 4.3, \( \lim_{N \to \infty} z(N) = z_m < \lim_{\beta \to \beta_c} \frac{1}{k_B \beta} \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} \) calculated in (4.24). The analytical calculation proving (4.30) is presented in Appendix B.2. A comparison between the values of \( Z(N) \) calculated numerically compared to the ones given by (4.30) in the order of \( N^{-1/3} \) is presented in Fig. 4.4.

![Figure 4.4: Graph for the value of \( z \), defined in (4.30), calculated numerically for \( N \) ranging from \( 10^2 \) to \( 10^8 \) (scattered blue points) and the approximation in order of \( N^{-1/3} \) (solid orange line) — meaning \( z(N) = z_m + \bar{z} N^{-1/3} \) as in (4.30). Figure previously published in [7].](image)

Other interesting properties can be observed from the study of Bose gases in finite \( N \). As commented in Sec. 4.3, \( c_v \) and its derivative are continuous for \( 0 < \eta < \frac{1}{2} \). The same
numerical investigation used in Figs. 4.1 - 4.3 can also illustrate an important difference in qualitative behaviour for this regime. In Fig. 4.5 the graphs for $c_v$ and $\left(\frac{\partial c_v}{\partial T}\right)_{N,\kappa}$ are presented for $\eta = 1/4$ — which is equivalent, per (4.11), to a interaction with $\tau = 8$ in a 2-dimensional gas. It is interesting to see that for $\eta = 1/4$, the specific heat at $\beta = 0$ is larger than for $\beta \rightarrow \beta_c$. In this case $c_v$ is increasing for small $\beta$ — in accordance to (4.17) — but as $\beta$ increases, it reduces smoothly — as expected from Table 4.1 — so no non-analytical behaviour is observed for $c_v$ or its first derivative at $\beta_c$.

4.5 Conclusions

The investigation presented in this chapter is a description of BE statistics that does not rely on the thermodynamic limit. This is made possible from the numerical calculation of $\xi(\beta, N)$ as the inverse of (4.12a). From this, all thermodynamical quantities can be written in terms of $\beta$ and $N$ allowing for a direct comparison to $\beta_c$.

The thermodynamical quantities that identify the BE condensation were calculated here. The fraction of particles in the ground state is calculated for arbitrary $\eta$ in terms of $\xi(\beta, N)$ in (4.25). Supplemented by the numerical inversion, graphs of this quantity for a gas trapped in a regular box ($\eta = 1/2$) and in a harmonic potential ($\eta = 2$) are presented in Fig. 4.1. Similarly the specific heat is calculated in (4.26) and the numerical results are presented in Fig. 4.2. Finally, the derivative of specific heat is calculated in (4.28) and presented in Fig. 4.3.

In all of these figures, the thermodynamical quantities were calculated for values of $N$ raging from $10^2$ to $10^7$ — in accordance to the numbers found in the experimental observation of BE condensation — where significant differences are observed in comparison to the calculations made in the thermodynamic limit. A summary for the convergence and continuity of these quantities in the thermodynamic limit is presented at Table 4.1. These graphs by themselves an important visualization of the role of the thermodynamic limit for the non analytical behaviour indicating the phase transition in BE gases, hence an important pedagogical tool for the study of phase transitions.

Particularly in Fig. 4.3 a fundamental difference in the qualitative behaviour of the specific heat derivative is observed. Considering particles in the ground state below critical
Figure 4.5: Graphs for $c_v(\beta, N)$ for $\eta = 0.25$ and $(\frac{\partial c_v}{\partial T})_{N,\kappa}$ in (4.28) of a Bose gas for $\eta = 1/4$. Unlike in the previous pictures, we see that both quantities are smooth for this value of $\eta$. Figure previously published in [7].
temperature, the minimum value of this quantity is smaller than the one given in the thermodynamic limit. This result is supported by analytical calculations (4.30) and it is found — both by numerical results in Fig. 4.4 and analytical calculations in (4.30) — that such minimum value scales with $N^{-1/3}$.

With the numerical inversion of (4.12a) established and available at [61], further studies on BE condensation for a broad range of quantum systems — whenever the density of states exponent $\eta$ can be identified — are now possible without relying on the thermodynamic limit nor on large $N$ approximations — the method presented here obtains calculations with arbitrary precision for any value of $\beta$ and $N$. 

78
CHAPTER 5

Entropic dynamics on Gibbs statistical manifolds

5.1 Introduction

The modern method of maximum entropy (MaxEnt) offers a fundamental understanding of probability and inference. It consists of updating probability distributions from priors to posteriors by maximizing the entropy functional under a set of constraints meant to represent the relevant information for the model at hand. In its most usual application the method selects distributions defined over a space of microstates meant to represent all the possible states of the system. Usually, the constraints are expected values of real functions of these microstates and each constraint generates a conjugate Lagrange multiplier. This results in the family of Gibbs distributions – also know as canonical distributions, Gibbs measures or the exponential family – which are the macrostates, meaning distributions that describe the statistical behavior of the system. The macrostates can either be parametrized by Lagrange multipliers or by the expected values.

Although MaxEnt is most known as a way to provide the foundations to equilibrium statistical mechanics, it is currently established as a universal framework for inference. Unsurprisingly, it has been applied to several fields such as economics, ecology, cellular biology, and opinion dynamics. As is the case with thermodynamics, all these applications are essentially static. MaxEnt has also been deployed to non-equilibrium statistical mechanics (see and subsequent literature in maximum caliber, e.g., ) but the dynamics is not intrinsic to the probabilities; it is induced by the underlying Hamiltonian dynamics of the molecules. For problems beyond physics there is a need for more general dynamical frameworks based on information theory.

In this chapter – based on the work published in – address such problems using entropic dynamics – a framework in which the laws of change for a given system are derived from MaxEnt. In the interpretation of probability theory as the logic system that deals with incomplete information, we assign unknown macrostates a probability distribution. When new information is presented MaxEnt is the appropriate tool to describe how these
distributions should be updated. This is applied by designing a dynamical process in which the probability of changes (or transitions) maximizes entropy under the constraints that give the appropriate information for the motion. The result is a stochastic process in which the parameter that takes the role of time is not given externally but arises from the accumulation of changes in the MaxEnt description. This is called entropic time is adjusted to best describe the relevant change in the dynamical process, and not necessarily bear a clear relation to physical time.

Entropic dynamics has been successful in reproducing dynamical models in physics such as quantum mechanics [62, 63], quantum field theory [64], and renormalization groups [1]. Also, it has been expanded recently into the fields of finance [131, 132] and neural networks [133]. Here we aim for different applications of entropic dynamics, creating the laws of dynamics driven by the natural geometry of the space of probability distributions.

The entropic dynamics developed here is not a form of nonequilibrium statistical mechanics. Although both describe changes for the macrostates, in nonequilibrium statistical mechanics they arise from general considerations over the dynamics for the microstates (which are, by themselves, derived from classical or quantum mechanics), while this entropic dynamics is completely agnostic of any microstate dynamics. This is particularly relevant because the use of methods once restricted to statistical physics in a broad range of scientific endeavours suggests the need for more general dynamical models. Unlike physics, most fields do not enjoy the same privilege of having a well established formalism for its microstates dynamics due to the complexity of the underlying process. Entropic dynamics can address it by giving a foundational principle from which laws of dynamics for macrostates, or its parameters, are derived.

This dynamics results in a stochastic process in which the temporal change of parameters is described by a diffusion – Fokker-Planck – equation. The drift, or the average change in the macrostates, will follow a continuous trajectory, whose derivative with respect to time, in a near flat region of the space, will be linear with respect to the gradient of entropy with symmetric (reciprocal) factors. This result is similar to the Onsager reciprocal relations [134] (ORR).

A diffusion process that reduces to a reciprocal relation in the linear regime is known in physics as the Onsager-Machlup process [135]. Unlike in EntDyn, the reciprocity found by
Onsager is a consequence of the time reversal symmetry in an underlying dynamics, hence the ORR are not trivially generalizable to dynamical processes beyond physics.

The chapter is organized as follows: Next section discusses the space of Gibbs distributions and its geometric properties; section 5.3 considers the ideas of entropic dynamics; section 5.4 tackles the difficulties associated with studying such a dynamics in the curved space of probability distributions; section 5.5 introduces the notion of entropic time; section 5.6 describes the evolution of the system in the form of a differential equation; section 5.7 will discuss the resemblance between entropic dynamics and ORR; finally section 5.8 presents different examples of applications to simple and well-known statistical models.

### 5.2 Entropic Dynamics

Having established the structure of this space of Gibbs distributions in the Sec. 2.3, we can use it to describe the changes for the macrostates. However, since the uncertainty here extends to both the microstate $x$ and the macrostates $\rho(x|A)$, we consider the change in the entire state of the system defined by the joint probability $P(x, A)$ given by

$$P(x, A) = \rho(x|A)P(A). \quad (5.1)$$

Where $\rho$ has the Gibbs form given in (2.58). That means we impose $P(x|A)$ to be canonical and the distribution $P(A)$ represents our lack of knowledge for the macrostates.$^{24}$

#### 5.2.1 Change happens

The goal here is to describe how change happens. If the state of a system is described by a set of expected values $A = \{A^1, A^2, ..., A^n\}$, we aim to derive laws that describe the change to another set $A' = \{A'^1, A'^2, ..., A'^n\}$. The incomplete information on these values, is codified in the probabilities for the values of $A$ and $A'$. It changes from a probability distribution $P(A)$ to another probability distribution $P(A')$. One can always write the marginalization process

$^{24}$Note that what we did in (5.1) is nothing more than assuming a probability distribution for the macrostates. This description is sometimes referred to as superstatistics [123].
\[ P(A') = \int dA' P(A', A) = \int dA' P(A'|A)P(A). \quad (5.2) \]

What entropic dynamics proposes is to choose the form of \( P(A'|A) \) by maximizing an entropy subject to whatever constraints properly reflect the relevant information about the motion. For example, in the entropy dynamics approach to quantum mechanics [63] the constraints lead to Brownian trajectories that conserve energy; in the entropic dynamics approach to finance [131] the constraints represent a change in prices and utilities. The investigation presented here proposes, under minimal assumptions, constraints for a dynamics of probabilities distributions on the statistical manifold of canonical distribution parametrized by the macrostates \( A \).

5.2.2 Transition probabilities

The core of entropic dynamics relies on assigning a conditional probability that maximizes entropy, given by

\[ S[P] = -\int dA' \int dx' P(x', A'|x, A) \log \left( \frac{P(x', A'|x, A)}{Q(x', A'|x, A)} \right), \quad (5.3) \]

Which is the entropy for the joint conditional probability that would describe a change in both the macrostates \( A \) and the microstates \( x \). At this stage their dynamics is described together, one of the constraints to be presented later will establish that as both \( x \) and \( A \) change, the probability of \( x' \) is already codified in \( A' \).

Although \( S \) in \((2.60)\) and \( S \) in \((5.3)\) are both entropies, in the information theory sense, they represent two very distinct statistical objects. The \( S \) in \((2.60)\) is familiar from statistical mechanics texts, which is the maximum entropy given an expected value constraint. While the \( S \) in \((5.3)\) is the entropy to be maximized so that we find the transition probability that better matches the information at hand, that means \( S \) is a tool to select the dynamics of the macrostates.
5.2.3 The prior

The $Q$ function in (5.3) is the prior. We want to extract a dynamics for the macrostates where $A$ is, a priori, uncorrelated with the microstate $x$ dynamics. This is achieved by a separable prior,

$$Q(x', A'| x, A) = Q(x'| A', x, A)Q(A'| x, A).$$

(5.4)

Here the a priori distribution for the transition of microstates should allow the microstate to change independently to any other microstate. That is achieved by selecting a prior transition in $x$ of the form $Q(x'| A', x, A) = q(x')$, the prior probability originally assigned for the MaxEnt (2.58) on the probability of microstates. Also the a priori probability for the updated macrostate should not depend on the old microstate, meaning $Q(A'| x, A) = Q(A'| A)$ leading to a prior of the form

$$Q(x', A'| x, A) = q(x')Q(A'| A).$$

(5.5)

Inspired by [136], we assume the a priori transition in $A$ to be continuous within the canonical manifold. Meaning the transition can be reduced to a series of short steps, $A'^i - A^i = \Delta A^i \to 0$, in which the relevant information is the infinitesimal distance element by FRIM, $dl^2 = g_{ij} dA^i dA^j$. The prior that accomplishes this, also obtained through MaxEnt and explained in the appendix C, is

$$Q(A'| A) \propto g^{1/2}(A') \exp \left[-\frac{1}{2\tau} g_{ij}(A) \Delta A^i \Delta A^j \right].$$

(5.6)

Where $g(A) = \det g_{ij}(A)$ and $1/2\tau$ the Lagrange multiplier for the continuous motion constraint. Note that this prior is only adequate for infinitesimal short steps as the exponent is only invariant in the limit $\tau \to 0$, so that $\Delta A$ converges in probability to 0. In Sec. 5.5, we are going to see how $\tau$ is related to a time interval.

5.2.4 The constraint

The information about the dynamics will be aligned to the design that the correlations between $x'$ and $x$ are only the ones from $A$. Hence, given $A'$, the probability for $x'$ does not depends explicitly on $A$ and $x$. The transition from a point $A$ to $A'$ is justified so that the
point \( A' \) corresponds to the probability distribution \( \rho(x'|A') \) in the same statistical manifold. The conditional probability can be separated as

\[
P(x', A'|x, A) = P(x'|A', x) P(A'|A, x)
\]

and we constraint the probability of \( x' \) to be canonical, obtaining

\[
P(x', A'|x, A) = \rho(x'|A') P(A'|A, x)
\]

### 5.2.5 Maximizing the entropy

Having specified the constraints we can maximize the entropy,

\[
S[P] = - \int dA' \int dx' \rho(x'|A') P(A'|A, x) \log \left( \frac{\rho(x'|A') P(A'|A, x)}{q(x') Q(A'|A)} \right).
\]

The terms for the integral in \( x' \) can be separated,

\[
S[P] = - \int dA' P(A'|A, x) \left[ \log \frac{P(A'|A, x)}{Q(A'|A)} + \int dx' \rho(x'|A') \log \frac{\rho(x'|A')}{q(x')} \right],
\]

and the second term inside the squared brackets is just the entropy for the canonical distribution as in (2.60). So, to find the conditional probability, we need to maximize

\[
S[P] = - \int dA' P(A'|A, x) \left[ \log \frac{P(A'|A, x)}{Q(A'|A)} + S(A') \right],
\]

which results in the distribution

\[
P(A'|A, x) = \frac{1}{\Xi} Q(A'|A) e^{S(A')} = \frac{1}{\Xi} g^{1/2}(A') \exp \left[ S(A') - \frac{1}{2\tau} g_{ij} \Delta A^i \Delta A^j \right],
\]

where \( Q \) is substituted by (5.6) and \( \Xi \) is a normalization factor. The conditioned probability of interest, \( P(A'|A) \) is found directly as \( P(A'|A, x) \) since it does not depend on \( x \). This is not surprising since neither the prior nor the constraints indicate any correlation between \( A' \) and \( x \).

The exponent in (5.12) has a quadratic term, as discussed when presenting the prior (5.6), the transition from \( A \) to \( A' \) has to be an arbitrarily small continuous change. This
allows for a linear approximation of $S$, making so that the exponential factor is quadratic in $\Delta A$

$$P(A'|A) = \frac{g^{1/2}(A')}{Z} \exp \left[ \frac{\partial S}{\partial A^i} \Delta A^i - \frac{1}{2\tau} g_{ij} \Delta A^i \Delta A^j \right].$$  \hspace{1cm} (5.13)

Where $e^{S(A)}$ was absorbed in the normalization factor $Z$. This is the transition probability selected from maximum entropy (5.3). However some mathematical difficulties arise from the fact that (5.13) is defined over a curved manifold. We are going to explore these mathematical issues and their consequences to motion in the following section.

5.3 Transition on Curved Spaces.

Since the statistical manifold is a curved space, we must understand how the transition probabilities (5.13) behave under a change of coordinates. As (5.6) and (5.13) require arbitrarily small step, we wish to express the transition probability, as well as quantities derived from it, calculated up to the order of $\tau$. The quadratic term, $\Delta A^i \Delta A^j$, in (5.13) is in the same order as $\tau$. Therefore, even in the limit $\tau \to 0$, the transition will be affected by curvature effects.

Having the transition probability written in a set of coordinates $A$, we change coordinated by transforming the distribution – as discussed in section 2.3.1.2 – and define an invariant object $p$ from

$$P(A) = g^{1/2}(A) p(A),$$ \hspace{1cm} (5.14)

that gives us the full information for the probability distribution.

5.3.1 Normal coordinates

It is convenient to write (5.13) in normal coordinates (NC) at $A$ – labeled with Greek letter indexes ($\mu, \nu$). In this coordinate system $g_{\mu\nu} = \delta_{\mu\nu}$ leading to $g(A') = 1$, and (5.13) becomes

$$P(A'|A) = g^{1/2}(A')p(A'|A) = \frac{1}{Z(A)} \exp \left[ \frac{\partial S}{\partial A^\mu} \Delta A^\mu - \frac{1}{2\tau} \delta_{\mu\nu} \Delta A^\mu \Delta A^\nu \right].$$  \hspace{1cm} (5.15)
The transition probability in NC is a Gaussian that can be explicit written by completing squares in the exponent terms one obtains

\[ P(A'|A) = \frac{1}{Z(A)} \exp \left[ -\frac{1}{2\tau} \delta_{\mu\nu} (\Delta A^\mu - V^\mu)(\Delta A^\nu - V^\nu) \right], \tag{5.16} \]

where

\[ V^\mu = \tau \delta^{\mu\nu} \frac{\partial S}{\partial A^\nu}. \tag{5.17} \]

In this description, the displacement \( \Delta A^\mu \) is expressed as a sum of an expected drift \( \langle \Delta A^\mu \rangle = V^\mu \) and a fluctuation \( \Delta w^\mu = \Delta A^\mu - \langle \Delta A^\mu \rangle \). This leads us to write the Gaussian transition in terms of \( \Delta w \)

\[ P(A'|A) = \frac{1}{Z(A)} \exp \left[ -\frac{1}{2\tau} \delta_{\mu\nu} \Delta w^\mu \Delta w^\nu \right], \tag{5.18} \]

which yields

\[ \langle \Delta w^\mu \Delta w^\nu \rangle = \tau \delta^{\mu\nu} \text{ and } \langle \Delta A^\mu \Delta A^\nu \rangle = \tau \delta^{\mu\nu} + o(\tau). \tag{5.19} \]

### 5.3.2 Changing to original coordinates

In order to transform back into the original coordinates we expand \( \Delta A^\mu \) in terms of \( \Delta A^i \) (see [136, 137]) calculated up to the order of \( \tau \).

\[ \Delta A^\mu = \frac{\partial A^\mu}{\partial A^i} \Delta A^i + \frac{1}{2} \frac{\partial^2 A^\mu}{\partial A^i \partial A^k} \Delta A^i \Delta A^k + o(\tau). \tag{5.20} \]

For convenience, we are going to write the first two moments of the motion in \( A \) as \( V^i = \langle \Delta A^i \rangle \) and \( W^{ij} = \langle \Delta A^i \Delta A^j \rangle \). Equation (5.20) shows that \( V^i \), does not transform contravariantly when expanded in the orders of \( \tau \). Also, at this point, we have no reason to believe that \( W^{ij} \) would transform 2-contravariantly. Because of this, as presented in [137], we will introduce auxiliary objects \( \tilde{V} \) and \( \tilde{W} \) that coincide with \( V \) and \( W \) in NC and transform as

\[ \tilde{V}^i = \frac{\partial A^i}{\partial A^\mu} V^\mu \text{ and } \tilde{W}^{ij} = \frac{\partial A^i}{\partial A^\mu} \frac{\partial A^j}{\partial A^\nu} W^{\mu\nu}, \tag{5.21} \]

respectively.
Writing (5.20) and isolating $\Delta A^i$ up to the order $\tau$ we obtain
\[
\Delta A^i = \frac{\partial A^i}{\partial A^\mu} \Delta A^\mu - \frac{1}{2} \frac{\partial A^i}{\partial A^\mu} \frac{\partial^2 A^\mu}{\partial A^j \partial A^k} \Delta A^j \Delta A^k ,
\] (5.22)

taking the expected value, we obtain
\[
\tilde{V}^i = V^i + \frac{1}{2} \Gamma^i_{jk} W^{jk} ,
\] (5.23)

where $\Gamma^i_{jk}$ are the Christoffel symbols. This allows us to compute $\tilde{V}^i$ from (5.17) and $\tilde{W}^{ij}$ from (5.19)
\[
\tilde{V}^i = \tau g^{ij} \frac{\partial S}{\partial A^j} \quad \text{and} \quad \tilde{W}^{ij} = \tau g^{ij} + o(\tau) .
\] (5.24)

From here, we can find the correlations in the original coordinates:
\[
\tilde{W}^{ij} = W^{ij} + \frac{1}{2} \Gamma^i_{lm} \langle \Delta A^l \Delta A^m \Delta A^j \Delta A^k \rangle + \frac{1}{2} \Gamma^j_{op} \langle \Delta A^i \Delta A^o \Delta A^p \rangle + \frac{1}{4} \Gamma^i_{lm} \Gamma^j_{op} \langle \Delta A^l \Delta A^m \Delta A^o \Delta A^p \rangle .
\] (5.25)

Observing the results of (5.17) and (5.19), we see that the expected values of product of three or more, such as $\langle \Delta A^i \Delta A^j \Delta A^k \rangle$, are of order higher than $\tau^{3/2}$. This means that, to order $\tau$, (5.25) implies that the second moments transform 2-contravariantly as
\[
\tilde{W}^{ij} = W^{ij} + o(\tau) = \tau g^{ij} + o(\tau) .
\] (5.26)

Then, from (5.23) and (5.24)
\[
V^i = \tau g^{ij} \frac{\partial S}{\partial A^j} - \frac{\tau}{2} \Gamma^i_{jk} g^{jk} + o(\tau) .
\] (5.27)

Therefore the first two moments, (5.17) and (5.19), in natural coordinates are
\[
\langle \Delta A^i \rangle = \tau g^{ij} \frac{\partial S}{\partial A^j} - \frac{\tau}{2} \Gamma^i , \quad \text{and} \quad \langle \Delta A^i \Delta A^j \rangle = \tau g^{ij} ,
\] (5.28)

where $\Gamma^i = \Gamma^i_{jk} g^{jk}$.

Note that we used several words such as ‘transitions’, ‘short step’, ‘continuous’ and ‘dynamics’ without any established notion of time. In the following section we will talk...
about time not as an external parameter, but as an emergent parameter from the maximum
entropy transition (5.13) and its moments (5.28).

5.4 Entropic Time

A challenge in our endeavor is: How to make changes accumulate? Our “laws of change”
should allow for a change from $A$ to $A'$, and then a change to a sequential third value $A''$.

5.4.1 Introducing time

To solve the impasse, we should go back to our set constraint (5.8). $A'$ and $A$ are ele-
ments of the same manifold, therefore $P(A')$ and $P(A)$ are two probability distributions over
the same space. Our established solution (see [62]) is to introduce a “book-keeping” param-
eter $t$, so that it would distinguish the said distributions as labelled by different parameters,
i.e. $P(A') = P_t(A')$ and $P(A) = P_t(A)$.

In this formalism we must call these different labels as a description of the system at
particular instants $t$ and $t'$. This allows us to call $P(A'|A)$ a transition probability, since
(5.2) becomes

$$P_t(A') = P(A') = \int dA P_{\Delta t}(A'|A)P_t(A)$$

(5.29)

where $\Delta t = t' - t$ is a duration.

As the system changes from $A$ to $A'$ and then to $A''$. The probability $P(A'')$ will be
constructed from $P(A')$, not explicitly dependent on $P(A)$. That means (5.29) represents
a Markovian process – conditioned to $P_t(A)$ the “future” $P_{t''}(A), t'' > t'$ is independent of
the “past” $P_t(A)$. It is important to notice that under this formalism (5.29) is not used to
show that the a process is Markovian in an existing time, but rather the concept of time and
dynamics developed here makes it Markovian by design.

It is also important to notice that the parameter $t$ presented here is not necessarily the
physical time. Our parameter $t$, which we call entropic time, is an epistemic well-ordered
parameter in which the dynamics are defined.$^{25}$

$^{25}$It is relevant to mention that, in the entropic dynamics approach to quantum mechanics [63, 64], entropic
time is the physical time.
5.4.2 The entropic arrow of time

It is important to note that the marginalization process in (5.2) could also lead to

\[ P(A) = \int dA' P(A|A')P(A') , \]  

(5.30)

where the conditional probabilities are related by Bayes Theorem,

\[ P(A|A') = \frac{P(A)}{P(A')} P(A'|A) , \]  

(5.31)

showing that a change “forward” will not happen the same way as a change “backwards” unless the system is in some form of stationary state, \( P(A) = P(A') \). Another way to present this is that probability theory alone gives no intrinsic distinction of the change “forward” and “backward”. The fact that we assigned the change “forward” by MaxEnt makes it so that, in general, the change “backward” is not an entropy maximum. Therefore, the preferential direction of timely flow arises from entropic dynamics naturally.

5.4.3 Calibrating the clock

In order to connect the entropic time to the transition probability, one needs to introduce the duration \( \Delta t \) in respect to the motion that makes (5.13) simple. The time interval will be chosen in terms of the fluctuations in (5.28) as

\[ \tau = \eta \Delta t , \]  

(5.32)

where \( \eta \) is a constant that fixed the units appropriately. In that description the parameter \( \tau \) that first appeared in the prior (5.6) takes the role of time. This particular choice is made so that for a short transition we have

\[ \langle \Delta A^i \Delta A^j \rangle = \eta \Delta t g^{ij} . \]  

(5.33)

That means, under the time intervals defined by (5.32) the entropic time is a measure of the systems’ fluctuation. Rather than having the changes in the system to be presented in terms of given time intervals (or an external clock), here the system is its own clock.
The moments in (5.28) can be written up to the order of $\Delta t$ as
\[
\frac{\langle \Delta A^i \rangle}{\Delta t} = \eta g^{ij} \frac{\partial S}{\partial A^j} - \frac{\eta}{2} \Gamma^i, \quad \text{and} \quad \frac{\langle \Delta A^i \Delta A^j \rangle}{\Delta t} = \eta g^{ij}.
\] (5.34)

The computed $\langle \Delta A^i \rangle$ suggests a connection to established results of fluctuation theory that will be explored in section 5.6. Before this, with the concept of time established, it is convenient to write the motion of $A$ as a differential equation.

5.5 Diffusion and the Fokker-Planck equation

Our goal of designing the dynamics from entropic methods is accomplished. The entropic dynamics equation of evolution is written in integral form as a Chapman-Kolmogorov equation (5.29) with a transition probability given by (5.12). In this section we will conveniently rewrite it in the differential form. The computed drift $\langle \Delta A^i \rangle$ and the fluctuation $\langle \Delta A^i \Delta A^j \rangle$ in (5.34) describe the dynamical process as a smooth diffusion [137]. Therefore, for a short transition, it is possible to write the evolution of $P_t(A)$, as a Fokker-Planck (diffusion) equation,
\[
\frac{\partial}{\partial t} P = -\frac{\partial}{\partial A^i} \left( P v^i \right),
\] (5.35)
where
\[
v^i = \eta g^{ij} \frac{\partial S}{\partial A^j} - \frac{\eta}{2} g^{ij} \frac{\partial}{\partial A^j} \left( \log \frac{P}{g^{1/2}} \right).
\] (5.36)

The derivation of (5.35) and (5.36) takes into account the fact that the space in which the diffusion happens is curved and is given in appendix B. In equation (5.36) we see that the current velocity $v^i$ consists of two components. The first term is the drift velocity guided by the entropy gradient and the second term is an osmotic velocity, that is a term that is driven by differences in probability density. The examples presented in section 8 will show how these terms interact and the dynamical properties derived from each.

5.5.1 Derivatives and divergence

Since both the entropy $S$ and the invariant density $p$ defined in (5.14) are scalars, the velocity defined in (5.36) is a contravariant vector. However, (5.35) is not a manifestly invariant equation. To check its consistency, it is convenient to write it in terms of the
\[
\frac{\partial}{\partial t}p = -\frac{1}{g^{1/2}} \frac{\partial}{\partial A^i} \left( g^{1/2} \, p v^i \right). 
\] (5.37)

In the right hand side we recognize the covariant divergence of the contravariant vector \(pv^i\), which can be written in the manifestly covariant form

\[
\frac{\partial}{\partial t}p = -D_i(pv^i), 
\] (5.38)

where \(D_i\) is the covariant derivative — see section 2.3.1.3. Having the dynamics fully described we can now study its consequences in the two following sections.

### 5.6 Specific Consequences: entropic dynamics and reciprocal relations

From the perspective of irreversible processes in thermodynamics one might recognize \(\Delta A\) in (5.34) to mimic the form of Onsager reciprocal relations (ORR), and with an additional term that can be interpreted as a correction for the fact that the motion of \(A^i\) has to account for probing curvature. This section will explore the consequences of it.

Assume the state of a thermodynamical system can be fully described by a finite number of real parameters \(\xi = \{\xi^1, \xi^2, ..., \xi^n\}\) and that thermodynamical entropy is a function of \(\xi\). In the neighbourhood of an equilibrium value \(\xi_0\), the rate of change of those parameters in time is assumed to be linear with respect to the gradient of entropy, meaning

\[
\frac{d\xi^i}{dt} = \gamma^{ij} \frac{\partial S}{\partial \xi^j}, 
\] (5.39)

and the terms of such linear transformation \(\gamma\), known as kinetic coefficients, are unspecified functions of \(\xi\). ORR state that the kinetic coefficients are symmetric, \(\gamma^{ij} = \gamma^{ji}\).

The change in time for \(\xi\) presented in (5.39) is fundamentally different from entropic dynamics. Firstly, it supposes a fully deterministic motion for the macrostates, while entropic dynamics creates a stochastic process that does not determine the values of the macrostates but rather their probability distribution. Second, the entropic dynamics presented here was only developed in the coordinates given by the expected values \(A^i\) and the only other coordinates considered are the Lagrange multipliers \(\lambda_i\). That is due to the fact that, from
the MaxEnt (2.55) application, $A^i$ are the natural variables for $S$.

The similarity between (5.34) and (5.39) could lead one to identify $\xi^i$ as $A^i$ in a flat space. That means, ORR would describe the trajectory for the expected values $\langle A^i \rangle$ and the unknown coefficients would be the unit corrected terms of the metric, $\gamma^{ij} = \eta g^{ij}$, that are clearly symmetric. However such a relationship is not direct. In the appendix C.3 we will present the derivation for ORR. From there, we can see that this identification would be equivalent as if the trajectories emerged from a macrostate rates of change were given by

$$
\frac{dA^i}{dt} = \eta (A^i_0 - A^i),
$$

(5.40)

that means, each macrostate value would move towards the equilibrium value $A_0$, with a change rate directly proportional to how far from the equilibrium they are.

Although (5.40) would make a trivial connection between entropic dynamics and nonequilibrium statistical mechanics, one might find the rates of change described by it to be an unrealistic model. In a thermodynamical example we can not expect macroscopic parameters, such as internal energy and number of particles, to evolve independently. We acknowledge that by stating that the dynamics developed here is the simplest possible dynamics developed from entropic considerations on the subject matter of canonical distributions.

If one knows that two variables are correlated and that information was not included in the analysis, the results will, likely, not give a good description of the system of interest. But neither MaxEnt nor entropic dynamics are to blame. In that case, it would be appropriate to develop a different dynamics using additional constraints that induce dynamical correlations between those macrostates. This would go beyond the scope of the present thesis, but the resultant reciprocal relations are an example of successful applications of entropic methods into dynamical systems in general and out of equilibrium statistical mechanics in particular.

That said, entropic dynamics offers a systematic method to find dynamics aligned with fundamental concepts of probability and statistics, while Onsager’s approach is based solely on calculus considerations around a supposed fixed point. Another way to point this out is to say that ORR is based on a non-statistical understanding of thermodynamics, while entropic dynamics is inspired by Jaynes’ information theory understanding of statistical physics.

The most important difference between the two methods is that, as previously stated,
entropic dynamics is completely agnostic of any microstate dynamics, unlike ORR that assumes a subdynamics that is time reversible. Also, since (5.1) is a completely general way to describe the probabilities for the microstates, the entropic dynamics developed in the previous section is applicable in any state of the system, not only near the equilibrium.

5.7 Examples

We established the entropic dynamics by finding the transition probability (5.13), presenting it as a differential equation in (5.35) and (5.36) and presenting it as invariant equation (5.38). We want to show some examples on how it would be applied and what are the results achieved. Our goal, as of now is not to search for realistic models, but to search for simple models which are both mathematically simple and general enough so it can give insight on how to use the formalism.

We will be particularly interested in two properties: the drift velocity,

\[ v^i_D = \eta g^{ij} \frac{\partial S}{\partial A_j}, \tag{5.41} \]

which is the first term of \( v^i \) in (5.36), and the static states \( v^i = 0 \), which are a particular subset of the dynamical system’s equilibrium \( \partial_i P = 0 \). These are obtained from (5.36) as

\[ v^i = 0 \Rightarrow \frac{\partial S}{\partial A^i} - \frac{1}{2} \frac{\partial}{\partial A^i} \log \left( \frac{P}{g^{1/2}} \right) = 0 \tag{5.42} \]

allowing one to write the unbounded stationary probability

\[ P(A) = \kappa g^{1/2}(A) \exp[2S(A)] \tag{5.43} \]

where \( \kappa \) is a normalization constant. This result shows that the invariant stationary probability density is \( p(A) = \kappa \exp[2S(A)] \).

5.7.1 General Gaussian Manifold

The space defined by the mean values and correlations of a random variable (or microstate) is the space of Gaussian distributions, which is an example of a canonical distribu-
Here we consider the dynamics of a two dimensional spherically symmetric Gaussian

\[
\rho(x \mid A) = \frac{1}{2\pi\sigma(A)^2} \exp\left(-\frac{1}{2\sigma(A)^2} \sum_{i=1}^{2} (x^i - A^i)^2\right)
\]  (5.44)

embedded in space defined by a mean dependent variance \(\sigma(A) = \sigma(A^1, A^2)\).

The space of Gaussians with a uniform variance, \(\sigma(A) = \text{constant}\), is flat and the dynamics turn out to be a rather trivial spherically symmetric diffusion. Choosing the variance to be non-uniform yields richer and more interesting dynamics. Because this example is pursued for purely illustrative purposes, we restrict to two dimensions and spherically symmetric Gaussians. The generalization is immediate.

The Fisher-Rao Information metric for a Gaussian distribution is found, using (2.80) – full derivation in [95], to be

\[
d\ell^2 = 4 \frac{\sigma^2}{\delta_{ij}} (d\sigma)^2 + \delta_{ij} dA^i dA^j
\]  (5.45)

so that, using \(d\sigma = \frac{\partial \sigma}{\partial A^i} dA^i\), the induced metric from \(d\ell^2 = g_{ij} dA^i dA^j\) is

\[
g_{ij} = \frac{1}{\sigma^2} \left(4 \frac{\partial \sigma}{\partial A^i} \frac{\partial \sigma}{\partial A^j} + \delta_{ij}\right).
\]  (5.46)

### 5.7.2 Gaussian submanifold around an entropy maximum

Since, entropy and the variance have a monotonic relation, as mentioned previously, in order to simulate an entropy maximum we can set

\[
\sigma(A) = \exp\left(-\frac{(A^1)^2 + (A^2)^2}{4}\right)
\]  (5.47)

which leads to a metric

\[
g_{ij} = \begin{bmatrix}
(A^1)^2 + \sigma^{-2} & A^1 A^2 \\
A^1 A^2 & (A^2)^2 + \sigma^{-2}
\end{bmatrix}
\]  (5.48)

such that equal volumes are measured by

\[
g^{1/2} = \sqrt{[(A^1)^2 + (A^2)^2] \sigma^{-2} + \sigma^{-4}}
\]  (5.49)
Figure 5.1: Drift velocity field for the entropy maximum example. Figure previously published in [3].

allowing one to write the drift velocity (Fig. 5.1)

\[ v_d^1 = -\frac{A^1\sigma^{-2}}{g} \quad \text{and} \quad v_d^2 = -\frac{A^2\sigma^{-2}}{g} \]

and unbounded stationary probability (Fig. 5.2)

\[ P(A) = 4\pi^2 g^{1/2}\sigma^{-4} \]

The static distribution results from the dynamical equilibrium between two opposite tendencies. One is the drift velocity field that drives the distribution along the entropy gradient towards the entropy maximum at the origin. The other is the osmotic diffusive force that we identified earlier as the ED analogue of Fick’s law. This osmotic force drives the distribution against the direction of the probability gradient and prevents it from becoming infinitely concentrated at the origin. At equilibrium, the cancellation between these two opposing forces results in the Gaussian distribution (5.51).
5.7.3 2-Simplex Manifold

Here we discuss an example with discrete microstates. The macrostate coordinates, being expected values, are continuous variables. Our subject matter will be a three-state system, \( x \in \{1, 2, 3\} \), such as, for example, a 3-sided die. The statistical manifold is the 2-dimensional simplex and the natural coordinates are the probabilities themselves,

\[
S_2 = \left\{ \rho(x) \mid \rho(x) \geq 0, \sum_{x=1}^{3} \rho(x) = 1 \right\}.
\]  

(5.52)

The distributions on the 2-simplex are Gibbs distributions defined by the sufficient statistics of functions

\[
a^i(x) = \delta_x^i \text{ so that } A^i = \langle a^i \rangle = \rho(i).
\]  

(5.53)

The entropy relative to the uniform discrete measure is

\[
S = -\sum_{i=1}^{3} \rho(i) \log(\rho(i)) = -\sum_{i=1}^{3} A^i \log(A^i),
\]  

(5.54)

and the information metric is given by

\[
g_{ij} = \sum_{k=1}^{3} \rho_k \frac{\partial \log(\rho_k)}{\partial A^i} \frac{\partial \log(\rho_k)}{\partial A^j}.
\]  

(5.55)

The 2-simplex arises naturally from probability theory due to normalization when one identifies the macrostate of interest to be the probabilities themselves. The choice of sufficient statistics \([5.53]\) implies that the manifold is a two-dimensional surface since, due to the normalization, one can write \( A^3 = 1 - A^1 - A^2 \). We will use the the tuple \((A^1, A^2)\) as our coordinates and \( A^3 \) as a function of them. In this scenario, one finds a metric tensor

\[
g_{ij} = \begin{bmatrix}
\frac{1}{A^3 + \frac{1}{A^1}} & \frac{1}{A^3} \\
\frac{1}{A^3} & \frac{1}{A^3 + \frac{1}{A^2}}
\end{bmatrix},
\]  

(5.56)
Figure 5.3: Drift velocity field for the 2-simplex in (5.58). Figure previously published in [3].

Figure 5.4: Static invariant stationary probability for the three-state system. Figure previously published in [3].

which induces the volume element

\[
g^{1/2} = \sqrt{\frac{1}{A^1 A^2 A^3}}.
\]

(5.57)

As is well known, the simplex is characterized by a constant curvature \( R = 1/2 \); the 2-simplex is the positive octant of a sphere. From (5.41) the drift velocity (Fig. 5.3) is

\[
v_D^1 = A^1 \left[ A^2 \log \left( \frac{A^2}{A^3} \right) + (A^1 - 1) \log \left( \frac{A^1}{A^3} \right) \right]
\]

\[
v_D^2 = A^2 \left[ A^1 \log \left( \frac{A^1}{A^3} \right) + (A^2 - 1) \log \left( \frac{A^2}{A^3} \right) \right],
\]

(5.58)

Also, the static probability is

\[
P(A) \propto g^{1/2} \prod_{i=1}^{3} (A^i)^{-2A^i}.
\]

(5.59)

We note from the determinant of the metric that the static probability (5.59) diverges at the boundary of the 2-simplex. This reflects the fact that a 2-state system (say, \( i = 1, 2 \)) is easily distinguishable from a 3-state system (\( i = 1, 2, 3 \)). Indeed, a single datum \( i = 3 \) will tell us that we are dealing with a 3-state system.

\[\text{26} \text{The ternary plots in Figures 5.3 and 5.58 were created using python-ternary library [138.]}\]
On the other hand, we can see (Fig. 5.4) that this divergence is not present in the invariant stationary probability (5.14).

As in the Gaussian case discussed in the previous section, the static equilibrium results from the cancellation of two opposing forces: the entropic force along the drift velocity field towards the center of the simplex is cancelled by the osmotic diffusive force away from the center.

5.8 Conclusions

The dynamics presented here is a result of a simple inference process leading to a diffusion process written in the differential form in (5.35) and (5.36). This dynamics on a general space of probability distributions relies solely on calculating the macrostate entropy $S(A)$ in (2.60), which can be computed with the same techniques for calculating entropy in statistical mechanics, and the metric $g_{ij}$ that can be calculated using (2.118) and (2.120). This rather simple dynamics is able to provide results analogous to ORR from geometric considerations, one could interpret this as indicative that some of the methods of fluctuation theory are not restricted to thermodynamics and might be extended to all fields for which MaxEnt is applied.

The dynamics was applied to two general spaces of probability distributions. A submanifold of the space of two-dimensional Gaussians and the space of probability distributions for a 3-state system (2-simplex). In each of these we were able to provide insights on how their dynamics work by calculating and plotting the drift velocity (5.41) and the stationary states (5.43).

Naturally this is a very simplistic dynamics as it only assumes continuous motion in a particular family of probability distributions written in the exponential form. However, considering the increasing relevance of MaxEnt for fields beyond physics, is important to point out that dynamics on the space of Gibbs distributions are in order. Entropic dynamics provides a framework for which many dynamical processes can be derived. If the dynamics presented here fails to identify some relevant real-world dynamical process, this can, in principle, be corrected by applying the relevant information as new constraints in the process of maximizing entropy to find $P(A'|A)$. 98
CHAPTER 6
Entropic dynamics of networks

6.1 Introduction

A graph is a set of nodes and a set of links connecting a pair of nodes. The number of links ending in a particular node is referred as the degree of such node. In network science, meaning is attached to a graph, as its nodes represent entities and its links represent interactions between entities (for example networks where the links represent publication coauthorships between scientists – nodes – [139], links representing associations between gene/disease [140] or scientific concepts [141]).

Gibbs distributions have been studied in the context of random graphs by Park and Newman [34], leading to an extensive investigation of MaxEnt applications in network science [35, 67, 68, 69, 70, 36, 37]. In this plethora of investigations, many models are proposed as different choices of the sample space – e.g. simple graphs, weighted graphs – and sufficient statistics – functions defined over the possible graphs – e.g. total connectivity (the total number of links in a graph), node degrees sequences (the degree of each node in the graph), and average nearest neighbour connectivity (the average degree of the direct neighbors of each node). However, considering these MaxEnt procedures, when one chooses a function as sufficient statistics it does not mean that a precise number is known for their expected values. As an example, MaxEnt obtains the correct distribution of link placement for these scale-free networks when one chooses the degree of each node as sufficient statistics, the power law behavior for the node degrees is fitted posteriorly with data. However, as explained by Radicchi et al. [37], this MaxEnt model alone does not justify why in some networks the node degrees are highly heterogeneous – meaning, they have long deviations from the first moment, as in a power law. Therefore an external model for sampling degree sequences is needed.

The issue described is a statement of the fact that, despite its generality, MaxEnt cannot tell by itself what constraints are relevant to the specific problem. The choice of constraints are justified by the fact that they work in practice, that is, they lead to a model
that accurately describes the system of interest. For example, in physics [76] we can assume 
that the microscopic world follows a conservative (Hamiltonian) dynamic, leading to expected 
value constraint of conserved quantities. Ultimately, unavoidable scientific labor is necessary 
to understand which constraints correctly implements the information one has about the 
system of interest.

In the previous chapter, an entropic formalism for dynamics in a space of Gibbs distribu-
tions was presented. The dynamics developed there relies on the concepts of information 
geometry\(^\text{27}\). The dynamics obtained is a diffusion process in the Gibbs statistical manifold 
- space of Gibbs distributions parametrized by the expected values and endowed with a 
Riemannian metric from information geometry.

A widely known example for dynamics of networks considers a preferential attachment 
mechanism – meaning, a mechanism through which new links are added to the network and 
the probability that a new link has an end in one of specific nodes is proportional to the 
degree of that node, see e.g [65 145 146] – leading to scale-free networks, where the degree 
distribution follows a power law. On the other hand, there has been literature challenging 
reported scale-free networks [66] and power laws in general [147]. These argue that one can 
not verify the power law behavior in real world networks under strong statistical validation, 
which indicates that scale-free networks are expected from a highly idealized processes and 
further dynamical models accounting for the peculiarities of a particular system are in order 
[148 149].

Our goal with the present chapter – related to the investigation published in [4] – is to 
show how EntDyn can provide a systematic way to derive dynamics of network ensembles. 
As an example, we apply the EntDyn developed in the previous chapter (and [3]) to the space 
of Gibbs distributions of graphs obtained after choosing the node degrees as constrains. We 
compare the steady state distributions obtained from EntDyn to the distributions found in 
real-world networks [66]. These results are not based on an underlying dynamics with par-
ticular assumptions, rather they are a consequence of the information geometry of networks 
ensembles. We bring into attention that although the dynamics developed here is simple, we 
also comment on how the framework provided by EntDyn is flexible enough so that further

\(^{27}\)Incidentally, information geometry has been applied to generate measures for complexity, see e.g. [142] 
[143 144 106].
constraints can be implements to account for the information available about the dynamical process.

In the following section, we will present the random graphs model used and the maximum entropy distributions and information geometry derived from it. In section 6.3, we review the entropic dynamics presented in the previous chapter (and [3]) on the context of random graphs obtaining a differential equation for the dynamics of networks. In section 6.4, we find the steady-states of the differential equation and argue on how the power law behavior emerges from the dynamics.

6.2 Gibbs distributions of graphs

In this section, we will establish the random graph model for the present chapter and obtain the Gibbs distribution and the metric tensor for its information geometric structure. A graph is defined by a set of nodes (or vertices) \( V \) and a set of links (or edges) \( E = \{\varepsilon_{\mu}\} \). Each link connects two nodes, thus, we will write \( \varepsilon_{\mu} = (i,j) \), where \( i,j \in \{1,2,\ldots,|V|\} \) are elements of an enumeration for the set of nodes \( V \). For the scope of the present chapter, we treat graphs in a general manner without attributing any information related to what the random graph may represent. Because of this, the constraint defined here and the dynamical assumptions in the next section will not assume a particular application.

6.2.1 MaxEnt of graphs

At this stage we will assign, through MaxEnt, a probability distribution \( \rho(G) \) for each graph (microstate) \( G = (V,E) \). Inspired by [37] — although similar descriptions have been proposed before e.g. [34, 67] — we will suppose a graph with \( N = |V| \) nodes and \( L = |E| \) links, and constraints on the number of links (also referred to as degree or connectivity) of each node \( i \).

To obtain the appropriate distribution \( \rho(G) \), in the context of MaxEnt, we ought to maximize the functional

\[
S[\rho|q] = -\sum_{G} \rho(G) \log \frac{\rho(G)}{q(G)} = -\sum_{E} \rho(E) \log \rho(E) ,
\]

(6.1)
where \( q \) is a prior distribution. The functional \( S[\rho|q] \) in (6.1) is known as Kullback-Leibler (KL) \(^{28}\) entropy, reducing to Shannon entropy when \( q \) is uniform. The last equality in (6.1) holds as we assume that we are inferring over an already known number of nodes \( N \) and uniform prior, meaning the ends of the links are not determined, but the total number of links is. Each link is treated independently under the same constraints of node connectivity. Since Shannon entropy is additive – meaning that for independent subsystems the joint entropy is the sum of entropies calculated for each subsystem – and preserves subsystems independence (see Sec 2.2.1) – meaning if two subsystems are independent in the prior and the constrains do not induce correlations then the posterior (distribution that maximizes entropy) will also be independent for each subsystem. Therefore, with separate constraints for each link, (6.1) reduces to

\[
S[\rho|q] = -\sum_{\epsilon} \left( \prod_{\mu} \rho(\epsilon_\mu) \right) \log \left( \prod_{\mu} \rho(\epsilon_\mu) \right) = -\sum_{\mu} \sum_{\epsilon_\mu} \rho(\epsilon_\mu) \log(\rho(\epsilon_\mu)) .
\]  

(6.2)

If the same constraint is applied to each link, MaxEnt results in distributions of the same form for each \( \epsilon_\mu \). Because of this, we can treat each link separately and write, for simplicity, \( \rho(\epsilon) = \rho(\epsilon_\mu) \) for every \( \mu \). So that (6.2) can be written as

\[
S[\rho|q] = L \, s[\rho] , \quad \text{where} \quad s[\rho] = -\sum_{\epsilon} \rho(\epsilon) \log \rho(\epsilon) ,
\]  

(6.3)

and \( s \) can be referred to as the entropy per link. Thus maximizing entropy \( S \) for the graph \( G \) is equivalent to maximizing \( s \) for a link \( \epsilon \).

To implement, in our MaxEnt procedure, that the relevant information is the degree of each node we define the functions \( a^i(\epsilon = (j,m)) = \frac{1}{2} (\delta_j^i + \delta_m^i) \) – where \( \delta_j^i \) refers to the Kronecker delta\(^{29}\) – as our sufficient statistics. Leading to the expected value constraints

\[
\langle a^i(\epsilon) \rangle = \sum_{j,m} \rho(\epsilon = (j,m)) \left( \frac{\delta_j^i + \delta_m^i}{2} \right) = \frac{k_i^i}{2L} = A^i ,
\]  

(6.4)

\(^{28}\)Even though the network problem might make one expect to obtain power laws it does not mean we should skew away from KL entropy. It has been widely reported that other functionals proposed to replace KL entropy, such as Renyi’s or Tsallis’, induce correlations not existing in the prior or constraints \(^{128, 150, 151, 2}\) and therefore lead to inconsistent statistics .

\(^{29}\)That means, \( \delta_i^i = 1 \) and \( \delta_j^i = 0 \) if \( i \neq j \). Equivalently \( \delta_{ii} = 1 \) and \( \delta_{ij} = 0 \) if \( i \neq j \).
where $k^i$ is the expected degree of each node $i$. The $2L$ factor is included so that the expected values $A^i$ sum to unity, since by construction, the sum of degrees is twice the number of links. The function that maximizes (6.3) under (6.4) and normalization is the Gibbs distribution

$$\rho(i,j|\lambda) = \frac{1}{Z} \exp \left( - \sum_m \lambda_m a^m(i,j) \right) = \frac{1}{Z} e^{-\frac{1}{2} \lambda_i - \frac{1}{2} \lambda_j},$$

(6.5)

where $Z$ is the normalization factor

$$Z = \sum_{i,j} e^{-\frac{1}{2} \lambda_i - \frac{1}{2} \lambda_j} = \left( \sum_i e^{-\frac{1}{2} \lambda_i} \right)^2,$$

(6.6)

and $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_N\}$ is the set of Lagrange multipliers dual to the expected values $A = \{A^1, A^2, \ldots, A^N\}$. In (6.5), we have the probability that a link $\varepsilon$ connects the nodes $i,j$ given the set of Lagrange multipliers. However, (6.4) indicates that they can also be parametrized by the expected values $A$. The two sets of parameters are related by

$$A^i = -\frac{1}{Z} \frac{\partial Z}{\partial \lambda_i} = \frac{e^{-\frac{1}{2} \lambda_i}}{\sum_j e^{-\frac{1}{2} \lambda_j}}.$$

(6.7)

Equating the previous result with (6.4) we obtain $k^i = e^{-\frac{1}{2} \lambda_i}$ allowing us to write

$$\rho(\varepsilon = (i,j)|A) = \frac{k^i k^j}{(2L)^2} = A^i A^j.$$

(6.8)

That is, we can interpret $A^i$ as the probability for which a specific link $\varepsilon$ has the node $i$ in one of its ends. For reasons that will be presented later in our investigation, it is also useful to calculate the node entropy at its maximum as a function of the expected values, meaning

$$s(A) = s[\rho(i,j|A)] = -2 \sum_i A^i \log A^i,$$

(6.9)

the last equality is found by substituting (6.8) into (6.3).

Since we can parametrize the space of probability distributions by expected values $A^i$ we will use those as coordinates when assigning the geometry to this space in the following subsection.
6.2.2 Information Geometry

Our present goal is to assign a Riemannian geometric structure to the distributions defined in (6.8). That means the space of Gibbs distributions is uniquely defined by the values of $A$, and the distances obtained from $d\ell^2 = \sum_{i,j} g_{ij} dA^i dA^j$ are a measure of distinguishability between the neighbouring distributions $\rho(i, j|A)$ and $\rho(i, j|A + dA)$. The metric components $g_{ij}$ are given by the Fisher-Rao information metric (FRIM) \[84, 85]\[
\begin{align*}
\frac{\partial \log \rho(m, n|A)}{\partial A^i} \frac{\partial \log \rho(m, n|A)}{\partial A^j}.
\end{align*}
\]
(6.10)
For details on the choice of the metric see Sec. 2.3.

Before calculating the FRIM for the distributions defined in (6.8) is important to remember that $\sum_i A^i = 1$. We will express then the value related to the last node in the enumeration as a dependent variable $A^N = 1 - \sum_{i=1}^{N-1} A^i$. Since we are describing graphs with a fixed number of links $L$, the constraints defined in (6.4) have some level of redundancy, namely $A^N$ is automatically defined by the set of all others. Even though this does not interfere with the maximization process – MaxEnt is robust enough to properly deal with redundant information – this has to be taken into account when calculating the summations in (6.10).

Therefore the FRIM components for the probabilities obtained in (6.8) are

$$g_{ij} = 2 \delta_{ij} A^i + 2 A^N$$
(6.11)

It would be useful to have an expression valid for all indices, $i, j \in [1, N - 1]$. For it we use, as in \[95\], that $dA^N = - \sum_{i=1}^{N-1} dA^i$, the expression for infinitesimal distances then becomes

$$d\ell^2 = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \left( \frac{2 \delta_{ij}}{A^i} + \frac{2}{A^N} \right) dA^i dA^j = \sum_{i,j} \frac{2 \delta_{ij}}{A^i} dA^i dA^j.$$
(6.12)

Yielding then a much simpler and diagonal metric tensor

$$g_{ij} = \frac{2}{A^i} \delta_{ij}.$$
(6.13)
As it is a property of Gibbs distributions [95, 3] this metric tensor could also have been found as the Hessian of \( s(A) \) in (6.9). A diagonal result is consistent with the fact that, per (6.8), both nodes at the end of a link will be sampled independently with the same distribution.

Having calculated the metric for the Gibbs distributions of our graph model, we have all elements to define a dynamics on it in the following section.

### 6.3 Entropic dynamics of Gibbs distributions

Entropic dynamics is a formalism for which the laws of dynamics are derived from entropic methods of inference. For the scope of the present chapter we are going to evolve the parameters \( A \) representing a change for the probabilities for \( \varepsilon \) in (6.8). This is equivalent to have distributions from which the sequences of node degrees, \( k^i = 2LA^i \), are selected from. In this description, the probabilities of links can be recovered from

\[
P(\varepsilon, A) = P(A)\rho(\varepsilon|A) ,
\]

where \( \rho \) is defined in (6.8). The dynamical process will describe how the change from a set of parameters \( A \) – representing an instant of the system– evolves to a set of parameters \( A' \) for which the distribution for a later instant \( P(A') \) is assigned as

\[
P(A') = \int dA \ P(A'|A)P(A)
\]

EntDyn consists on finding the transition probability \( P(A'|A) \) through the methods of information theory. As done in our previous chapter, the dynamical process will rely on two assumptions: (i) the changes happens continuously\(^{30}\) which will determine the choice of prior and (ii) that the motion has to be restricted to the Gibbs distributions obtained from \( \rho(\varepsilon|A) \) in (6.8) which will determine our constraint. Beyond the scope of our investigation, different models can be generated by imposing additional constraints that implement other information that might be relevant to the dynamical process.

\(^{30}\)Continuous motion might not sound as a natural assumption in a discrete system, such as graphs, however even if a space is discrete, the set of probability distributions on it is continuous as are the expected values that parametrize it.
The entropy we need to maximize has to account for the joint change in the degrees of uncertainty in the graph connections $\varepsilon$ as well as the parameters $A$ which is represented by the distribution in (6.14). The transition from $A$ to $A'$ must also contain information about the transitions from $\varepsilon$ to a later link distribution $\varepsilon'$. Therefore, we must maximize entropy for the joint transition $P(\varepsilon', A' \mid \varepsilon, A)$, meaning

$$ S[P \mid Q] = - \sum_{\varepsilon'} \int dA' P(\varepsilon', A' \mid \varepsilon, A) \log \left( \frac{P(\varepsilon', A' \mid \varepsilon, A)}{Q(\varepsilon', A' \mid \varepsilon, A)} \right), \quad (6.16) $$

where $Q(\varepsilon', A' \mid \varepsilon, A)$ is the prior to be determined. We shall call $S$ the dynamical entropy to avoid confusion with the graph entropy in (6.1) and the entropy per node (6.3).

The prior that implements continuity for the motion on the statistical manifold but is otherwise uninformative is of the form

$$ Q(\varepsilon', A' \mid \varepsilon, A) = Q(\varepsilon' \mid \varepsilon, A, A') Q(A' \mid \varepsilon, A) \propto g^{1/2}(A') \exp \left( -\frac{1}{2\tau} g_{ij} \Delta A^i \Delta A^j \right), \quad (6.17) $$

as explained in chapter 5, where $\Delta A^i = A'^i - A^i$, $g = \det g_{ij}$, and $\tau$ is a parameter that will eventually take the role of a short time interval, since when $\tau \to 0$ leads to short steps, meaning $d\ell \to 0$.

The constraint that expresses the fact that the motion does not leave the space of Gibbs distributions defined in Section 2.1 is

$$ P(\varepsilon', A' \mid \varepsilon, A) = P(\varepsilon' \mid \varepsilon, A, A') P(A' \mid \varepsilon, A) = \rho(\varepsilon' \mid A') P(A' \mid \varepsilon, A). \quad (6.18) $$

That means the distribution for $\varepsilon'$ conditioned on $A'$ must be of the form (6.8). Note that per (6.18) the only factor still undetermined for the full transition probability is $P(A' \mid \varepsilon, A)$.

The result obtained when maximizing (6.16) with the prior (6.17) and under (6.18) is

$$ P(A' \mid \varepsilon, A) \propto g^{1/2}(A') \exp \left( s(A') - \frac{1}{2\tau} g_{ij} \Delta A^i \Delta A^j \right). \quad (6.19) $$

Note that it is independent of $\varepsilon$, which is not surprising since neither the prior nor the constraints assume any correlation between $A'$ and $\varepsilon$ hence $P(A' \mid A) = P(A' \mid \varepsilon, A)$. For short steps, $\tau \to 0$ therefore $\Delta A \to 0$, we can Taylor expand $s$ in terms of $\Delta A$, leading to the
transition probability of the form

\[ P(A'|A) = \frac{1}{Z(A)} g^{1/2}(A') \exp \left( \sum_i \frac{\partial s}{\partial A^i} \Delta A^i - \sum_{ij} \frac{1}{2\tau} g_{ij} \Delta A^i \Delta A^j \right), \tag{6.20} \]

where the normalization factor \( Z(A) \) absorbs the proportionality constant in (6.19) and \( e^{\psi(A)} \).

In section 5.3.2, we calculate the moments for this transitions up to order \( \tau \) obtaining

\[
\langle \Delta A^i \rangle = \tau \sum_{j,k} \left( g^{ij} \frac{\partial s}{\partial A^j} - \frac{1}{2} \Gamma^{i}_{jk} g^{jk} \right),
\]

\[
\langle \Delta A^i \Delta A^j \rangle = \tau g^{ij} \quad \text{and} \quad \langle \Delta A^i \Delta A^j \Delta A^k \rangle = o(\tau);
\]

where \( g^{ij} \) are the elements of the inverse matrix to \( g_{ij} \) and \( \Gamma^{i}_{jk} \) are the Christoffel symbols.

Equation (6.21) is the definition of a smooth diffusion \[137\] if we choose \( \tau \) as a time duration \( \Delta t \), which is equivalent to calibrating our time parameter in terms of the fluctuations \( \Delta t \approx \tau \propto \sum_{ij} g_{ij} \Delta A^i \Delta A^j \). That means, here the role of time emerges from emergent properties of the motion, up to a multiplying constant, time measures the fluctuations in \( A \). The system is its own clock. As explained in \[3\] this leads to the evolution as a Fokker-Planck equation

\[
\frac{\partial p}{\partial t} = -\frac{1}{g^{1/2}} \sum_i \frac{\partial}{\partial A^i} \left( g^{1/2} p v^i \right), \quad \text{where} \quad v^i = g^{ij} \frac{\partial}{\partial A^j} \left( s - \frac{1}{2} \log(p) \right), \tag{6.22} \]

and \( p \) is the invariant probability density \( p(A) \equiv \frac{P(A)}{\sqrt{g(A)}} \). If we substitute the graph entropy (6.9) and the metric (6.13) we obtain

\[
\frac{\partial p}{\partial t} = \sum_i \left( \left[ \frac{1}{2} \log A^i + \frac{3}{2} \right] p + \left[ A^i (\log A^i + 1) + \frac{1}{8} \right] \frac{\partial p}{\partial A^i} + A^i \frac{\partial^2 p}{\partial A^i \partial A^i} \right), \tag{6.23} \]

This establishes the dynamical equation for the graph model. In the following section we will focus on finding a steady-state \( \bar{p}(A) \) for (6.23).
6.4 Entropic steady state of graphs models

In order to find a steady state in (6.23) we note that that for $\frac{dp}{dt} = 0$ the equation is separable, meaning the solution can be written as a product of the same function, $p(a)$, for each argument $\tilde{p}(A = \{A^i\}) = \prod_i p(a = A^i)$. This implies that each term in the summation on (6.23) has to be zero and therefore $p(a)$ has to obey

$$\frac{a}{4} \frac{d^2 p}{da^2} + \left[ a(\log a + 1) + \frac{1}{8} \right] \frac{dp}{da} + \left[ \frac{1}{2} \log a + \frac{3}{2} \right] p = 0 . \tag{6.24}$$

In order to solve the above equation we make the substitution $y = \sqrt{8a}$, so that it transforms into

$$\frac{d^2 p}{dy^2} + y[f(y) - 2] \frac{dp}{dy} + f(y)p = 0 , \tag{6.25}$$

where $f(y) = 2\log y - \log 8 + 3$. This substitution is equivalent to write (6.22) under a change of coordinates $Y^i = \sqrt{8}A^i$, in which the metric (6.13) transforms into an Euclidean metric, $\sum_{ij} g_{ij} dA^i dA^j = \sum_{ij} \delta_{ij} dY^i dY^j$.

The range at which (6.25) is valid takes into account the fact that the maximum possible connectivity is the number of links, $k = L$, leading to a possible maximum value for $a = 1$ and $y = \sqrt{8}$, when self-connections are not ignored. However, Anand et. al. [69] argues that in order for the connectivity of each node to remain uncorrelated, a lower maximum connectivity should be considered. Inspired by their arguments we can set $y_{max} = \sqrt{25/2L}$, corresponding to $k_{max} = \sqrt{2L}$ and $a_{max} = 1/\sqrt{2L}$. Also, we can see that (6.25) diverges at $y = 0$ unless $p(y = 0) = 0$. Therefore, we'll consider $y \in (0, y_{max}]$.

Solving (6.25) is enough to obtain the steady-state values for $p(a)$ in (6.23) and therefore the degree distribution

$$P(a = \frac{k}{2L}) = \sqrt{\frac{2}{a}} p(y = \sqrt{8a}) , \tag{6.26}$$

where the square root factor comes from the information metric $\sqrt{g(a)} = \sqrt{\frac{2}{a}}$. We choose to solve (6.25) as an initial value problem (IVP) by making sure that the node connectivity remains uncorrelated, thus setting

$$p(y = y_{max}) = 0 \quad \text{and} \quad \frac{dp}{dy} \bigg|_{y=y_{max}} = dp_0 , \tag{6.27}$$

108
where \( y_{\text{max}} = \sqrt{\frac{25}{L}} \) was considered for \( L = 2^9, 2^{13} \) and \( 2^{30} \). The final result is later normalized, and the precise value of \( d_p \) is to be investigated.

The degree distribution \( P(a = \frac{k}{2L}) \) obtained from this method is presented in Fig. 6.1, where we see that, under normalization, the value of \( d_p \) does not influence the probability values. Furthermore, upon the rescaling \( \bar{a} = a/a_{\text{max}} \) and \( \bar{P} = P\sqrt{a_{\text{max}}} \), or similarly \( \bar{y} = y/y_{\text{max}} \), the number of links \( L \) does not alter the behaviour of the degree distribution, as seen in Fig. 6.2.

Another initial value condition we investigated was to consider every node in the graph to have at least one connection,

\[
p(y = 0) = 0 \quad \text{and} \quad \frac{dp}{dy}_{y=0} = d_p ,
\]

(6.28)

The integration runs until \( y = \sqrt{8} \), where \( a = 1 \), and then normalized. Similarly to the previous case, the value of \( d_p \) does not interfere with the degree distribution after normalization, as shown in Fig. 6.3.

Inspired by the distributions reported to be found in real-world networks \[66\], we fit the numerical results for \( P(a) \) to the fat-tailed distributions in Table 6.1. The range – region at which the distribution is valid – is chosen based on the values of \( a \) that minimize the root mean squared error (RMSE) between the functional form and the numerical results. The fact that for Fig. 6.2 the solutions start from above zero means that the degree distribution is considered from the minimal possible non-zero connectivity, \( k = 1 \). Also, the case where
Figure 6.2: Re-scaled degree distribution for the IVP (6.27) irrespective of the number of links $L$. The result fits well with a Weibull and Gamma, also known as power law with cutoff, distributions within most of the allowed range. Figure previously published in [4].

Figure 6.3: Degree distribution derived from the solution of the IVP (6.28). The value of $d_{p0}$ does not interfere with the degree distribution. The result fits well with a Weibull distribution for a network with many connections, $L > 10^4$. Figure previously published in [4].
Weibull($a; \lambda, k$) \[ \left( \frac{a}{k} \right)^{\lambda-1} e^{-\left( \frac{a}{k} \right)^{\lambda}} \] \[ a \in (0, 2 \cdot 10^{-1}] \text{ in Fig. 6.2} \] \[ \text{RMSE} = 0.118 \]

Gamma($a; \lambda, k$) \[ a^{-\lambda} e^{-ka} \] \[ a \in (0, 8 \cdot 10^{-1}] \text{ in Fig. 6.2} \] \[ \text{RMSE} = 0.166 \]

PLIEC($a; \lambda, k$) \[ a^{-\lambda} e^{-k/a} \] \[ a \in [6 \cdot 10^{-5}, 2 \cdot 10^{-2}] \text{ in Fig. 6.2} \] \[ \text{RMSE} = 0.155 \]

Power-Law($a; \lambda$) \[ a^{-\lambda} \] \[ a \in (0, 5 \cdot 10^{-3}] \text{ in Fig. 6.2} \] \[ \text{RMSE} = 0.385 \]

Table 6.1: Summary of the real-world degree distributions fit to the results of the IVPs (6.27) and (6.28). The chosen range takes into account the value of the root mean square error (RMSE), which is minimized relative to the entire range of $a$ in the plot. Table previously published in [4].

every node has at least one connection only fits with a real-world degree distribution for networks with $L > 10^4$ links. Note from (6.26) that the dynamics leads to a natural power law behavior as $a^{-0.5}$, which was found throughout most of Fig. 6.2 range.

6.5 Conclusions

We presented an entropic dynamics of graphs with a fixed number of nodes $N$ and connections $L$. This model leads to the Gibbs distribution (6.8) whose information metric is given by (6.13). When the dynamical assumption is that we evolve the parameters continuously and constrained in the statistical manifold we are lead into the Fokker-Plank equation (6.22). Steady-state solutions for two different IVPs are presented: The differential equation results for (6.27) are graphically represented in Fig. 6.1 and fitted for Weibull and Gamma distributions in Fig. 6.2. The resulting differential equation, solved for (6.28), was fitted for the Weibull distribution are presented in Fig 6.3 where distribution found in real-world networks [66] is only valid for graphs with many $L > 10^4$ links.

Our result is an information theory approach for dynamics of networks in which, under very general assumptions, the power law behavior emerges. Naturally, this work is not the single process for dynamics of networks. Rather, under this method, other random graph models can be studied and other constraints – instead of or in addition to (6.18) – can be implemented when maximizing the dynamical entropy $S$ as in (6.16). The investigation presented here demonstrates that it is possible to derive a dynamical process in complex systems using information theory.
A.1 On the low fugacity limit of curvature

In this appendix we will explicitly calculate all steps needed for finding the low fugacity limit for curvature in FD statistics (3.38) and BE statistics (3.39). For the purposes of our investigation we will assume that the value of $\eta$ is so that all of the following quantities do not diverge: $\Gamma(\eta + 1), \Gamma(\eta + 2)$, $\Gamma(\eta + 3)$, and $\Gamma(\eta + 4)$. This is observed in all models presented in Table 3.1, furthermore $\eta > -1$ is a sufficient condition for the assumption to hold true.

Using the first term expansion of polylogarithms (3.35) in the calculation of $A$ in (3.20) we obtain

$A(x, \eta) = f(\eta)x^2 + o(x^2), \quad (A.1)$

where $f(\eta) \doteq \Gamma(\eta + 3)\Gamma(\eta + 1) - \Gamma(\eta + 2)^2$. Equivalently for $A_c$ in (3.29) we obtain

$A_c(x, \eta) = f_c(\eta)x^2 + o(x^2), \quad (A.2)$

where $f_c(\eta) \doteq \Gamma(\eta + 3)$. Also, in the first term expansion of polylogarithms for $B$ in (3.23) we obtain

$B(x, \eta) = o(x^3), \quad (A.3)$

that means, $B$ vanishes in order $x^3$. This is not surprising since we can see in (3.23) that the first order expansion of polylogarithms turns $B$ into the determinant of a matrix with duplicate lines. However, since we are interested in calculating the limits of curvature in (3.38) and (3.39), $B$ needs to be compared to $A^2$, and by squaring (A.1) we see that the first non-vanishing term of $A^2$ is of order $x^4$. Hence, we need to compute $B$ at least up to order $x^4$.

If we use the second term expansion of the polylogarithm – meaning expanding (3.13)
as \( \text{Li}(x, \phi) = x + 2^{-\phi}x^2 + o(x^2) \) – we obtain

\[
\mathcal{B}(x, \eta) = h(\eta)x^4 + o(x^4),
\]

where

\[
h(\eta) \doteq \frac{1}{2\eta + 1} \left[ - \Gamma(\eta + 1)\Gamma(\eta + 2)\Gamma(\eta + 4) + \frac{3}{2} \Gamma(\eta + 1)\Gamma(\eta + 3)\Gamma(\eta + 3) - \frac{1}{2} \Gamma(\eta + 2)\Gamma(\eta + 2)\Gamma(\eta + 3) \right].
\]

(A.5)

Equivalently, for \( \mathcal{B}_c \) in (3.32) we obtain

\[
\mathcal{B}_c(x, \eta) = h_c(\eta)x^4 + o(x^4),
\]

where

\[
h_c(\eta) \doteq \frac{1}{2\eta + 1} \left[ \Gamma(\eta + 2)\Gamma(\eta + 4) - \frac{1}{2} \Gamma(\eta + 3)\Gamma(\eta + 3) \right] + 2 \left[ \Gamma(\eta + 3)\Gamma(\eta + 3) - \Gamma(\eta + 2)\Gamma(\eta + 4) \right].
\]

(A.7)

Hence, the curvature for FD statistics (3.21) in the low fugacity limit is given by

\[
\lim_{\xi \to 0} R_{FD} = \frac{\beta^{\eta+1}}{2\kappa} \lim_{x \to 0^+} \left[ \frac{\mathcal{B}(x, \eta)}{(\mathcal{A}(x, \eta))^2} \right] = \frac{h(\eta)}{(f(\eta))^2} \frac{\beta^{\eta+1}}{2\kappa},
\]

(A.8)

and the curvature for BE statistics (3.30) in the same limit is given by

\[
\lim_{\xi \to 0} R_{BE} = -\frac{\beta^{\eta+1}}{2\kappa} \lim_{x \to 0^+} \left[ \frac{\mathcal{B}(x, \eta) + \frac{\beta^{\eta+1}}{\kappa} \mathcal{B}_c(x, \eta)}{(\mathcal{A}(x, \eta) + \frac{\beta^{\eta+1}}{\kappa} \mathcal{A}_c(x, \eta))^2} \right]
\]

\[
= -\left[ \frac{h(\eta) + h_c(\eta)\frac{\beta^{\eta+1}}{\kappa}}{(f(\eta) + f_c(\eta)\frac{\beta^{\eta+1}}{\kappa})^2} \right] \frac{\beta^{\eta+1}}{2\kappa}.
\]

(A.9)

For the 3-dimensional gas in a box, \( \eta = \frac{1}{2} \), we can calculate \( h^{(1/2)} \approx 0.6921 \), \( h_c^{(1/2)} \approx -5.321 \), \( f^{(1/2)} \approx 1.178 \), and \( f_c^{(1/2)} \approx 3.323 \). Substituting these values (A.8) becomes (3.38) and (A.9) becomes (3.39) completing the calculation.
APPENDIX B
Appendixes to chapter 4

B.1 Calculations in the thermodynamic limit

This appendix will derive the thermodynamical quantities of interest – calculated for finite \(N\) in (4.25), (4.26) and (4.28) – reduce to the ones presented in the thermodynamic limit – respectively (4.16), (4.17) and (4.20).

As explained in Sec. 4.3 \(\beta < \beta_c\) it implies \(n_0 = 0\) in the thermodynamic limit. It follows directly that \(\frac{n_0}{N} = 0\), reducing to (4.16) for \(\beta < \beta_c\). From implicit differentiation of (4.12a), it follows that

\[
\frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} = \frac{1}{\beta} (\eta + 1) \frac{\text{Li}_{\eta+1}(\xi)}{\text{Li}_{\eta}(\xi)},
\]

which is equivalent to (4.27) in the thermodynamic limit for \(\beta < \beta_c\). Similarly, it follows from the second differentiation of (4.12a) that

\[
\frac{1}{\xi} \left( \frac{\partial^2 \xi}{\partial \beta^2} \right)_{N,\kappa} = -\frac{1}{\beta^2} \left[ (\eta + 2)(\eta + 1) \frac{\text{Li}_{\eta+1}(\xi)}{\text{Li}_{\eta}(\xi)} - (\eta + 1)^2 \frac{\text{Li}_{\eta+1}(\xi)}{\text{Li}_{\eta}(\xi)} 
- (\eta + 1)^2 \frac{(\text{Li}_{\eta+1}(\xi))^2}{(\text{Li}_{\eta}(\xi))^3} \right] 
\]

which is equivalent to (4.29) in the thermodynamic limit for \(\beta < \beta_c\). Substituting (B.1) into (4.26) one obtains (4.17) for \(\beta < \beta_c\) and substituting (B.1) and (B.2) into (4.28) one obtains (4.20) for \(\beta < \beta_c\), completing the calculation.

As \(\beta \geq \beta_c\) it implies \(\xi = 1\) in the thermodynamic limit. Substituting \(\beta_c\) in (4.15) into (4.25) becomes (4.16) for \(\beta \geq \beta_c\). Analogously, substituting \(\xi = 1\) into (4.27) and (4.29) it follows directly that \(\left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} = 0\). Therefore, (4.26) becomes (4.17) for \(\beta \geq \beta_c\) and from the direct differentiation of (4.17) one obtains (4.20) for \(\beta \geq \beta_c\), completing the calculation.
B.2 Minimum value of $(\frac{\partial c_v}{\partial T})_{N,\kappa}$

This appendix will derive (4.30) analytically by an expansion of (4.28), thus explaining the observance of values of $(\frac{\partial c_v}{\partial T})_{N,\kappa}$ for $\eta = 1/2$ smaller than those found in the thermodynamic limit in (4.20) – as presented from numerical calculations in Fig. 4.3. This is done by calculating $(\frac{\partial c_v}{\partial T})_{N,\kappa}$, finding its minimum in a large $N$ approximation.\(^{31}\)

Two variables will be important for this calculation. The first, $\beta^*$, is the argument to the minimum value of the quantity of interest — abscissa of the minimum values for each $N$ in Fig. 4.3 — meaning, $\beta^* = \arg \min_{\beta} \frac{1}{k_B^2} (\frac{\partial c_v}{\partial T})_{N,\kappa}$. The second, $\xi^*$ is defined as the fugacity at the minimum value of the quantity of interest, meaning $\xi^* = \xi(\beta^*, N)$. From this, two other variables can be constructed: the reduced inverse temperature at the minimum $\gamma^* = \gamma N - \psi$ and $\lambda_2^* = \log \xi^*$, whose notation $\lambda_2$ is inspired by it being the second Lagrange multiplier at the minimum, as explained in section 3.2.

If one assumes a scaling relation between $\beta^*$ and $\lambda_2^*$ of the form

$$\gamma^* = \bar{\gamma} N^{-\psi} \quad \text{and} \quad \lambda_2^* = \bar{\lambda}_2 N^{-\phi}, \quad \text{(B.3)}$$

in the leading order of $N$ — where $\bar{\gamma}$ and $\bar{\lambda}_2$ are constants and $\psi$ and $\phi$ are positive. Substituting those variables in (4.12a) and $\beta_c$ in (4.15) one obtains

$$N = N(1 + \gamma^*)^{-3/2} \frac{\text{Li}_{3/2}(e^{-\lambda_2^*})}{\zeta^{3/2}} + \frac{1}{e^{\lambda_2^*} - 1}. \quad \text{(B.4)}$$

Note that accounting for only the first term would yield the regular calculation in the thermodynamic limit — expressed previously in (4.18).

Using the scaling relations in (B.3), the series expansion for polylogarithms in (4.14) and $e^{\lambda_2^*} = 1 + \lambda_2^* + o(\lambda_2^*)$ one can rewrite (B.4) in the leading terms,

$$-\frac{3}{2} \bar{\gamma} N^{1-\psi} - 2 \frac{\Gamma(1/2)}{\zeta^{3/2}} \bar{\lambda}_2^{1/2} N^{1-\phi/2} + \bar{\lambda}_2^{-1} N^\phi = 0. \quad \text{(B.5)}$$

A result that depends on both $\beta^*$ and $\lambda_2^*$ requires that the first and at least one other term in (B.5) must contribute to the highest order in $N$. If only the first two terms contribute, the result would ignore the particles in the ground state, leading to the same results in

\(^{31}\)The calculation presented in this appendix was done in collaboration with D. Robbins.
the thermodynamic limit — equivalent to (4.18). If only the first and last term in (B.5) contribute, one would find \( \gamma = \frac{3}{2} \tilde{\lambda}^{-1} \). This result however is undesirable physically, as observed in Fig. 4.3 we can expect \( \beta^* < \beta_c \) and, consequentially, \( \gamma < 0 \); and for BE statistics one must have \( \xi^* \leq 1 \) implying \( \tilde{\lambda} > 0 \). Therefore, it follows that all terms in (B.5) must contribute to the highest order, accounting for these terms one obtains \( 1 - \psi = 1 - \frac{\phi}{2} = \phi \), hence \( \psi = \frac{1}{3} \) and \( \phi = \frac{2}{3} \) — later these values will be verified numerically.

In order to obtain the values of \( z(N) \) one needs to substitute \( \beta^* \) and \( \xi^* \) in (4.28). In order to do so, it is necessary to first substitute these values in \( \frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \) in (4.27) and \( \frac{1}{\xi} \left( \frac{\partial^2 \xi}{\partial \beta^2} \right)_{N,\kappa} \) in (4.29) as a expansion in terms of \( N \). The parameters \( \gamma \) and \( \tilde{\lambda} \) will later be identified by imposing \( \frac{\partial}{\partial \beta} \left[ \frac{1}{k_B \beta} \left( \frac{\partial^2 \xi}{\partial T^2} \right) \right]_{\beta = \beta^*} = 0 \). This will be done in the following subsections.

### B.2.1 Expanding \( \frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \)

One can expand the numerator of (4.27), \( Q_n \) as

\[
Q_n = \frac{1}{\beta_c} \frac{3}{2} N + o(N) ,
\]

(B.6)

where \( o \) stands for the smaller order notation, \( \lim_{N \to \infty} \frac{o(f(N))}{f(N)} = 0 \). Similarly, for the denominator of (4.27), \( Q_d \), is expanded as

\[
Q_d = a N^{1/3} (1 + b N^{-1/3}) + o(N) ,
\]

(B.7)

where

\[
a = \frac{1}{\bar{\lambda}^2} + \frac{\Gamma(1/2)}{\zeta(1/2)} \frac{1}{\bar{\lambda}^{1/2}} \quad \text{and} \quad b = \frac{1}{a} \left[ \frac{\zeta(1/2)}{\bar{\lambda}^{1/2}} - \frac{3}{2} \frac{\Gamma(1/2)}{\zeta(3/2)} \frac{\bar{\lambda}}{\bar{\lambda}^{1/2}} \right] .
\]

(B.8)

Therefore, using \( \frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} = \frac{Q_n}{Q_d} \) it follows that

\[
\frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa, \beta = \beta^*} = \frac{1}{\beta_c} \bar{q} N^{-1/3} + o\left(N^{-1/3}\right) ,
\]

(B.9)

where

\[
\bar{q} = \frac{3}{2a} .
\]

(B.10)
Note that the second term for $Q_d$ in (B.7) does not appear in (B.9). The importance of calculating the second term in $Q_d$ will be shown to be relevant when other quantities are calculated from it, as it will be done in the following subsection.

B.2.2 Expanding $\frac{1}{2} \left( \frac{\partial^2 \xi}{\partial \beta^2} \right)_{N, \kappa}$

One can expand the numerator of (4.29), $F_n$, as

$$F_n = -\frac{a}{\beta_c^2} r_m N^{4/3} + \frac{a}{\beta_c^2} \bar{r} N + o(N)$$  \hspace{1cm} (B.11)

where

$$r_m = -\frac{1}{a^3} \left[ \frac{9}{2} \frac{1}{\lambda_2^{3/2}} + \frac{3 \Gamma(5/2)}{2 \zeta(3/2)} \frac{1}{\lambda_2^{3/2}} \right]$$  \hspace{1cm} (B.12a)

$$\bar{r} = \frac{1}{a^3} \left[ \frac{15}{4} a^2 - \frac{9}{2} \frac{\Gamma(1/2)}{\zeta(3/2)} \frac{a}{\lambda_2^{1/2}} + \frac{9 \Gamma(5/2)}{2 \zeta(3/2)} \frac{b}{\lambda_2^{3/2}} + \frac{27}{2} \frac{b}{\lambda_2^3} + \frac{39 \Gamma(5/2)}{4} \frac{\bar{\gamma}}{\zeta(3/2)} \frac{1}{\lambda_2^{3/2}} 
+ \frac{45}{2} \frac{\bar{\gamma}}{\zeta(3/2)} - \frac{9}{2} \frac{\Gamma(-1/2)}{\zeta(3/2)} \frac{1}{\lambda_2^{5/2}} - \frac{3}{\zeta(3/2)^2} \frac{\Gamma(-1/2) \Gamma(5/2)}{\lambda_2} \right]$$  \hspace{1cm} (B.12b)

Note that the denominator of (4.29) is the same as $Q_d$, expanded in (B.7). Therefore it follows that

$$\left. \frac{1}{\xi} \frac{\partial^2 \xi}{\partial \beta^2} \right|_{\beta=\beta^*} = \frac{r_m}{\beta_c^2} + \frac{\bar{r}}{\beta_c^2} N^{-1/3} + o(N^{-1/3})$$  \hspace{1cm} (B.13)
B.2.3 Expanding \( \frac{1}{k_B \beta} \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} \)

The quantity of interest, \( \frac{1}{k_B \beta} \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} \) in (4.28), can be expressed by substituting \( \beta_c \) in (4.15) as

\[
\frac{1}{k_B^2 \beta} \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} = \frac{3}{2} \left( \frac{\beta}{\beta_c} \right) \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} \bigg|_{\beta = \beta^*} = 3 \left( \frac{\beta}{\beta_c} \right) \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} \bigg|_{\beta = \beta^*} = z_m + z N^{-1/3} + o(N^{-1/3}) .
\]

(B.14)

using the expression for polylogarithms (4.14) and the results of the previous subsections – (B.9) and (B.13) – this can be expanded as

\[
\frac{1}{k_B^2 \beta} \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} \bigg|_{\beta = \beta^*} = z_m + z N^{-1/3} + o(N^{-1/3}) .
\]

(B.15)

where

\[
z_m = \frac{3}{2} \left( \frac{1}{\zeta(3/2)} \right) \left[ r_m \zeta(3/2) + \frac{15}{4} \zeta(5/2) \right] \quad \text{and} \quad \bar{z} = \frac{3}{2} \left( \frac{1}{\zeta(3/2)} \right) \left[ q^2 \Gamma(1/2) \bar{\lambda}_2^{-1/2} - 3 q \zeta(3/2) + \bar{\zeta}(3/2) + \frac{1}{2} r_m \bar{\gamma} \zeta(3/2) + r_m \bar{\lambda}_2^{1/2} \Gamma(-1/2) - \frac{45}{8} \bar{\zeta}(5/2) \right] .
\]

(B.16a, b)

Thus obtaining the scaling of \( \frac{1}{k_B^2 \beta} \left( \frac{\partial c_v}{\partial T} \right)_{N, \kappa} \) expressed in (4.30). In order to complete the goals of this appendix, one needs to obtain the values of \( \bar{\gamma}, \bar{\lambda}_2, a, b, \bar{q}, r_m, \) and \( \bar{r} \) and substitute those in (B.16). This will be done in the next two subsections.
B.2.4 Obtaining \( \bar{\gamma} \) and \( \bar{\lambda}_2 \)

Substituting the values of \( \psi \) and \( \phi \) in (B.5), it follows that

\[
\frac{3}{2} \bar{\gamma} + 2 \frac{\Gamma(1/2)}{\zeta(3/2)} \bar{\lambda}_2^{1/2} - \bar{\lambda}_2^{-1} = 0.
\]  

(B.17)

by implicit derivation of the equation above one finds

\[
\frac{d\bar{\lambda}_2}{d\bar{\gamma}} = -\frac{3}{2} \left( \bar{\lambda}_2^{-2} + \frac{\Gamma(1/2)}{\zeta(3/2)} \bar{\lambda}_2^{-1/2} \right)^{-1},
\]  

(B.18a)

\[
\frac{d^2\bar{\lambda}_2}{d\bar{\gamma}^2} = -\frac{2}{3} \left( \frac{d\bar{\lambda}_2}{d\bar{\gamma}} \right)^3 \left( 2\bar{\lambda}_2^{-3} + \frac{\Gamma(3/2)}{\zeta(3/2)} \bar{\lambda}_2^{-3/2} \right),
\]  

(B.18b)

\[
\frac{d^3\bar{\lambda}_2}{d\bar{\gamma}^3} = -\frac{4}{9} \left( \frac{d\bar{\lambda}_2}{d\bar{\gamma}} \right)^5 \left( 6\bar{\lambda}_2^{-6} + \frac{\Gamma(5/2)}{\zeta(3/2)} \bar{\lambda}_2^{-9/2} \right).
\]  

(B.18c)

The minimum occurs when

\[
0 = 3 \frac{\beta_{c}^{-1}}{2 \zeta(3/2)} \left[ -\left( \frac{\beta}{\beta_{c}} \right)^{-3/2} \frac{5}{2} \left( \frac{3}{2} \right)^2 \text{Li}_{3/2}(\xi) + \left( \frac{\beta}{\beta_{c}} \right)^{-3/2} \frac{5}{2} \frac{3}{2} \xi \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \text{Li}_{3/2}(\xi) \right.
\]

\[
+ \left( \frac{\beta}{\beta_{c}} \right)^{-1/2} \frac{5}{2} \frac{\beta_{c}}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \text{Li}_{1/2}(\xi)
\]

\[
- \left( \frac{\beta}{\beta_{c}} \right)^{-1/2} \frac{5}{2} \frac{\beta_{c}}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)^2_{N,\kappa} \text{Li}_{1/2}(\xi)
\]

\[
+ \left( \frac{\beta}{\beta_{c}} \right)^{1/2} \frac{\beta_{c}}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa}^3 \left( 2\text{Li}_{3/2}(\xi) - 3\text{Li}_{1/2}(\xi) \right) + \text{Li}_{-1/2}(\xi))
\]

\[
- \left( \frac{\beta}{\beta_{c}} \right)^{1/2} \frac{3}{2} \frac{\beta_{c}}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \left( \frac{\beta_{c}}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)^2_{N,\kappa} \right) \text{Li}_{3/2}(\xi) - \text{Li}_{1/2}(\xi))
\]

\[
+ \left( \frac{\beta}{\beta_{c}} \right)^{1/2} \frac{\beta_{c}}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \text{Li}_{1/2}(\xi)
\].

(B.19)

In order to solve (B.19) one may have to expand \( \frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \) — as done for \( \frac{1}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N,\kappa} \) in Sec. B.2.1 and \( \frac{1}{\xi} \left( \frac{\partial^2 \xi}{\partial \beta^2} \right)_{N,\kappa} \) in Sec. B.2.2 However, a less laborious manner to perform this
calculation involves identifying from (B.3) that

$$\frac{\beta_c}{\xi} \left( \frac{\partial \xi}{\partial \beta} \right)_{N, \kappa} = - \frac{d \lambda_2}{d \gamma} = - \frac{d \bar{\lambda}_2}{d \bar{\gamma}} N^{-1/3} + o(N^{-1/3}) \ , \quad (B.20a)$$

$$\frac{\beta^2}{\xi} \left( \frac{\partial^2 \xi}{\partial \beta^2} \right)_{N, \kappa} = - \frac{d^2 \lambda_2}{d \gamma^2} = - \frac{d^2 \bar{\lambda}_2}{d \bar{\gamma}^2} + o(1) \ , \quad \text{and} \quad (B.20b)$$

$$\frac{\beta^3}{\xi} \left( \frac{\partial^3 \xi}{\partial \beta^3} \right)_{N, \kappa} = - \frac{d^3 \lambda_2}{d \gamma^3} = - \frac{d^3 \bar{\lambda}_2}{d \bar{\gamma}^3} N^{1/3} + o(N^{1/3}) \ . \quad (B.20c)$$

Later it will be shown that $\bar{q} = - \frac{d \bar{\lambda}_2}{d \bar{\gamma}}$, as expected from (B.9), and $r_m = - \frac{d^2 \bar{\lambda}_2}{d \bar{\gamma}^2}$, as expected from (B.13). \[32\]

Expanding (B.17) it follows that

$$0 = - \frac{d^3 \bar{\lambda}_2}{d \bar{\gamma}^3} N^{1/3} \zeta^{3/2} + o(N^{1/3}) \ . \quad (B.21)$$

Hence, the minimum condition implies $\frac{d^3 \bar{\lambda}_2}{d \bar{\gamma}^3} = 0$, which is equivalent, per (B.18c), to

$$\text{6} \bar{\lambda}_2^{-6} + \frac{\Gamma(5/2)}{\zeta^{5/2}} \bar{\lambda}_2^{-9/2} = 0 \implies \bar{\lambda}_2 = \left[ \frac{1 \Gamma(5/2)}{6 \zeta^{5/2}} \right]^{-2/3} \ , \quad (B.22)$$

and applying this into (B.17) it follows that

$$\bar{\gamma} = \frac{2}{3} \left[ \bar{\lambda}_2^{-1} - 2 \frac{\Gamma(1/2)}{\zeta^{3/2}} \bar{\lambda}_2^{1/2} \right] \ , \quad (B.23)$$

leading to the values

$$\bar{\lambda}_2 \approx 5.1804 \quad \text{and} \quad \bar{\gamma} \approx -1.9303 \ . \quad (B.24)$$

The numerical verification of (B.3) with these values of $\bar{\gamma}$ and $\bar{\lambda}_2$ is presented in Fig. B.1.

\[32\]Note that this does not invalidate the work done in Sec. B.2.1 and B.2.2 since (B.20b) does not obtain the second term in (B.13).
Figure B.1: Graph for the values of $\gamma^*$ (above) and $\lambda_2^*$ (below) calculated numerically for $N$ ranging from $10^2$ to $10^8$ (scattered blue points) and the approximation in order of $N^{-1/3}$ and $N^{-2/3}$ (solid orange line) respectively — meaning $\gamma^*(N) = \bar{\gamma}N^{-1/3}$ and $\lambda_2^*(N) = \bar{\lambda}_2N^{-2/3}$ as in (B.3), with $\bar{\lambda}_2$ and $\bar{\gamma}$ given by (B.24). Figure previously published in [7].
B.2.5 Obtaining $a, b, \bar{q}, r_m, \bar{r}, z_m, \text{ and } \bar{z}$

Substituting the values of $\bar{\gamma}$ and $\bar{\lambda}_2$ from (B.24) into (B.8), one obtains

$$a \approx 0.33536 \quad \text{and} \quad b \approx 0.90686 . \quad \text{(B.25)}$$

Sequentially applying these values in (B.10) yields

$$\bar{q} \approx 4.47284 . \quad \text{(B.26)}$$

Note that substituting the value of $\bar{\lambda}_2$ from (B.24) into (B.18a) implies $\bar{q} = -\frac{d\bar{\lambda}_2}{d\bar{\gamma}}$ in accordance to (B.20a). Similarly, applying (B.26), (B.25), and (B.24) into (B.12) yields

$$r_m \approx -2.5746 \quad \text{and} \quad \bar{r} \approx -6.1656 . \quad \text{(B.27)}$$

Note that substituting the value of $\bar{\lambda}_2$ from (B.24) into (B.18b) implies $r_m = -\frac{d^2\bar{\lambda}_2}{d\bar{\gamma}^2}$ in accordance to (B.20b). Finally, substituting (B.24), (B.26), and (B.27) into (B.16) one obtains $z_m$ and $\bar{z}$ as in (4.31) completing the calculation.
C.1 Obtaining the prior

In this appendix we derive the prior transition probability from $A$ to $A'$ seen in (5.6). This is achieved by maximizing the entropy

$$S[Q] = \int dA' \ Q(A'|x,A) \log \left( \frac{Q(A'|x,A)}{R(A'|x,A)} \right),$$

(C.1)

where $R(A'|x,A)$, the prior for (C.1) would be an earlier stage of information for the systems’ dynamics. The posterior of (C.1), $Q(A'|x,A)$, becomes the prior in (5.3). At this stage $A$ could evolve into any $A'$ and the only assumption is that the assigned prior for (C.1) would give equal probabilities for equal volumes. That is achieved by a prior proportional to the volume element $R(A'|x,A) \propto g^{1/2}(A)$, where $g(A) = \det g_{ij}(A)$. There is no need to address normalization of $R$ since it will not effect in the posterior.

The constraint is so that the motion will be isotropic and continuous on the manifold. This will be imposed by

$$\int dA' \ Q(A'|x,A) \ g_{ij} \Delta A^i \Delta A^j = K.$$

(C.2)

where $K$ is a small quantity that eventually will tend to zero. This is so, due to the expected value of $g_{ij} \Delta A^i \Delta A^j$ being invariant only in the limit for short steps $\Delta A^i \to 0$.

The result of maximizing (C.1) under (C.2) and normalization is

$$Q(A'|x,A) \propto g^{1/2}(A') \exp \left( -\alpha \ g_{ij} \Delta A^i \Delta A^j \right),$$

(C.3)

where $\alpha$ is the Lagrange multiplier associated to (C.2). As the result requires $K \to 0$ to make it geometrically invariant, the conjugated Lagrange multiplier should equally be allowed to be taken to infinity. This allows us to define $\tau = 1/\alpha$, such that the short step limit leads to $\tau \to 0$. 

123
Note that, since no motion in $x$ and no correlation between $x$ and $A'$ is induced by the constraints, the result does not depend on the previous microstate $x$, $Q(A'|x,A) = Q(A'|A)$.

### C.2 Derivation of the Fokker-Planck equation

The goal of this appendix is to show that for a dynamics that is a smooth diffusion\(^{33}\) in a curved space, can be written as a Fokker-Planck equation and to obtain its velocity (5.36) from the moments for the motion (5.34). In order to do so, it is convenient us to define the drift velocity from

$$b^i = \lim_{\Delta t \to 0} \frac{\langle \Delta A^i \rangle}{\Delta t} = g^{ij} \frac{\partial S}{\partial A^j} - \frac{1}{2} \Gamma^i . \tag{C.4}$$

First let us analyze the change in a smooth integrable function $f(A)$ as the system transitions from $A$ to $A'$,

$$\Delta f(A) = \frac{\partial f}{\partial A^i} \Delta A^i + \frac{1}{2} \frac{\partial^2 f}{\partial A^i \partial A^j} \Delta A^i \Delta A^j + o(\Delta t) , \tag{C.5}$$

where a cubic term, $\Delta A^i \Delta A^j \Delta A^k$, would be $o(\Delta t)$. In a smooth diffusion we can take an expected value of (C.5) with respect to $P(A'|x,A)$.

$$\langle \Delta f(A) \rangle = \int dA' P(A'|x,A) (f(A') - f(A)) = \left( b^i \frac{\partial}{\partial A^i} + \frac{1}{2} g^{ij} \frac{\partial^2}{\partial A^i \partial A^j} \right) f(A) \Delta t . \tag{C.6}$$

which (C.6) can be further averaged in $P(A)$

$$\int dA' P(A') f(A') - \int dA P(A) f(A) = \int dA P(A) \left( b^i \frac{\partial}{\partial A^i} + \frac{1}{2} g^{ij} \frac{\partial^2}{\partial A^i \partial A^j} \right) f(A) \Delta t . \tag{C.7}$$

As established in section 5, $P(A)$ and $P(A')$ are distributions at the instants $t$ and $t'$ respectively.

$$\int dA \left( \frac{P_{t'}(A) - P_t(A)}{\Delta t} \right) f(A) = \int dA P(A) \left( b^i \frac{\partial}{\partial A^i} + \frac{1}{2} g^{ij} \frac{\partial^2}{\partial A^i \partial A^j} \right) f(A) , \tag{C.8}$$

\(^{33}\)Here smooth diffusion means, as defined by [137], a stochastic process in which the first two moments are calculated to the order of $\Delta t$, $\langle \Delta A^i \rangle = b^i \Delta t$, $\langle \Delta A^i \Delta A^j \rangle = \eta g^{ij} \Delta t$ and $\langle \Delta A^i \Delta A^j \Delta A^k \rangle = 0$.
which can be partially integrated in the limit of small steps

\[ \int \! dA \left( \frac{\partial P(A)}{\partial t} \right) f(A) = \int \! dA \left( -\frac{\partial}{\partial A^i} (b^i P(A)) + \frac{1}{2} \frac{\partial^2}{\partial A^i \partial A^j} (g^{ij} P(A)) \right) f(A) . \]  

(C.9)

Due to the generality of \( f \) as test function, we identify the integrands,

\[ \frac{\partial}{\partial t} P(A) = -\frac{\partial}{\partial A^i} \left( b^i P(A) - \frac{1}{2} \frac{\partial}{\partial A^j} (g^{ij} P(A)) \right) , \]  

(C.10)

and substitute \( b^i \) (C.4) for general coordinates,

\[ \frac{\partial}{\partial t} P(A) = -\frac{\partial}{\partial A^i} \left( g^{ij} \frac{\partial S}{\partial A^j} P(A) - \frac{1}{2} \Gamma^i P(A) - \frac{1}{2} \frac{\partial}{\partial A^j} ( \log P g^{1/2} ) P(A) \right) , \]  

(C.11)

and the contracted Christoffel symbols can be substituted in the identity

\[ \Gamma^i = -\frac{1}{g^{1/2}} \frac{\partial}{\partial A^j} (g^{1/2} g^{ij}) = -\frac{\partial g^{ij}}{\partial A^j} - g^{ij} \frac{\partial \log g^{1/2}}{\partial A^i} \]  

(C.12)

obtaining

\[ \frac{\partial}{\partial t} P(A) = -\frac{\partial}{\partial A^i} \left( g^{ij} \frac{\partial S}{\partial A^j} - \frac{1}{2} g^{ij} \frac{\partial}{\partial A^j} \left( \log \frac{P(A)}{g^{1/2}} \right) \right) P(A) , \]  

(C.13)

The result is a Fokker-Planck equation that is usefully written in the continuity form

\[ \frac{\partial}{\partial t} P = -\frac{\partial}{\partial A^i} \left( P v^i \right) , \]  

(C.14)

where

\[ v^i = g^{ij} \frac{\partial S}{\partial A^j} - \frac{1}{2} g^{ij} \frac{\partial}{\partial A^j} \left( \log \frac{P}{g^{1/2}} \right) , \]  

(C.15)

completing the derivation.

C.3 On the derivation of Onsager reciprocal relations

In this appendix we will comment on how the Onsager reciprocal relations (ORR) are derived in a general nonequilibrium statistical mechanics and how they compare to the entropic dynamics developed in the main text. This derivation is largely inspired by the
derivation provided by Landau and Lifshitz [10] translated into covariant notation.

This describes a thermodynamic entropy as a function of a set of parameters $\xi$ that change in time, we will study such a change as a fluctuation around an equilibrium value $\xi_0$. This equilibrium value has to both be a local maxima of entropy and a fixed point for the dynamics of $\xi$. That means

$$\frac{\partial \xi^i}{\partial t} \bigg|_{\xi = \xi_0} = 0 \quad \text{and} \quad \frac{\partial S}{\partial \xi^i} \bigg|_{\xi = \xi_0} = 0,$$

(C.16)

entropy has also to be concave with respect to each $\xi^i$ to guarantee that this will describe a maximum.

In order to show the relationship in (5.39) we will Taylor expand $d\xi^i/dt$ around the equilibrium point

$$\frac{d\xi^i}{dt} = L^i_j (\xi^j - \xi^j_0),$$

(C.17)

where $L^i_j$ would be identifiable as the gradient of the said rates of change,

$$L^i_j = \left. \frac{\partial}{\partial \xi^j} \frac{d\xi^i}{dt} \right|_{\xi = \xi_0},$$

(C.18)

the terms of L are not computable. They would need a knowledge of the deterministic dynamics for each value $\xi^i$ so that we would be able to know the gradients of their rates of change, which is not guaranteed in a statistical description of any system.

A second step towards ORR is to Taylor expand the gradients of entropy similarly near the equilibrium

$$\frac{\partial S}{\partial \xi^i} = \beta_{ij} (\xi^j - \xi^j_0),$$

(C.19)

where $\beta$ is the Hessian of the entropy

$$\beta_{ij} = \left. \frac{\partial^2 S}{\partial \xi^i \partial \xi^j} \right|_{\xi = \xi_0}.$$

(C.20)

Unlike $L$, $\beta$ is computable through standard methods and it is symmetric. We can directly

34It is interesting to note that Landau and Lifshitz mention Onsager in noncovariant notation. Despite their knowledge of differential geometry, they were unaware of the geometric structure behind probability distributions.
put (C.17) and (C.19) together achieving

\[ \frac{d\xi_i}{dt} = L_j^i [\beta^{-1}]^{jk} \frac{\partial S}{\partial \xi_k}, \]  

(C.21)

by direct comparison to (5.39) we see that

\[ \gamma_{ij} = L_k^i [\beta^{-1}]^{kj}. \]  

(C.22)

The proof that \( \gamma \) is symmetric relies on time reversal symmetry. In particular, it means that as the values of \( \xi \) change in time correlations between them should evolve in time so that:

\[ \langle \xi^i(0)\xi^j(t) \rangle = \langle \xi^i(-t)\xi^j(0) \rangle. \]  

(C.23)

The full proof need not be reproduced here (see [10, 134]). However, we note that the assumption of time-reversal symmetry, which applies to the microstate dynamics of atoms and molecules, will not necessarily hold for systems beyond physics. It may therefore be surprising that entropic dynamics, which does not rely on time reversibility, implies similar symmetric reciprocity relations.

If we restrict (C.19) to the motion of only the expected values \( \xi^i = A^i \) to mach the dynamics presented before we have

\[ \beta_{ij} = \frac{\partial^2 S}{\partial A^i \partial A^j} \bigg|_{A=A_0} = - g_{ij}(A_0), \]  

(C.24)

the choice of coordinates makes so that we can clearly see that the expansion presented in ORR is intrinsically geometrical. Also in these coordinates we can write the kinetic coefficients as

\[ \gamma_{ij} = -L_k^i g^{kj}(A_0), \]  

(C.25)

note that \( g^{ik}g_{kj} = \delta^i_j \).

So, for that choice of parameters, we can write (5.39) as:

\[ \frac{dA^i}{dt} = -L_k^i g^{kj}(A_0) \frac{\partial S}{\partial A^j}. \]  

(C.26)
One might naively assume that the relation between ED and ORR would be to identify the motion of the expected values of entropic dynamics without taking into account the first-order motion will probe into the curvature naturally given by FRIM. If that were to be correct a direct comparison to (5.34) would yield

\[ L_j^i = -\delta_j^i, \]  

(C.27)

that is better written by checking directly into how we defined \( L \) in (C.17),

\[ \frac{d\xi^i}{dt} \propto (\xi_o^i - \xi^i). \]  

(C.28)
BIBLIOGRAPHY


[38] Fisher, R.A.: On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 222(594-604) (1922) 309


[40] Pitman, E.J.G.: Sufficient statistics and intrinsic accuracy. 32(4) (1936) 567


136


