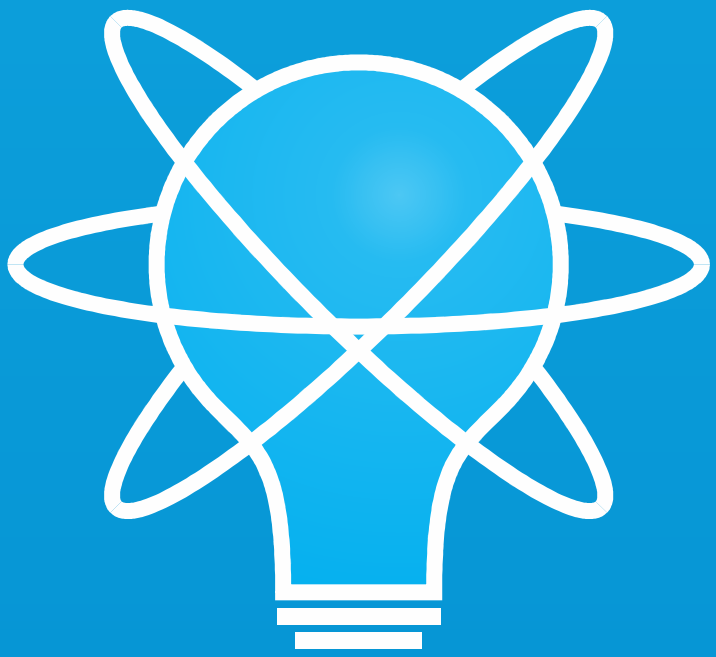
Assumptions of Physics



Gabriele Carcassi
Christine A. Aidala

Gabriele Carcassi, Christine A. Aidala

Assumptions of physics

Working DRAFT for Ver 3.1 - February 7, 2026

ASSUMPTIONSOFPHYSICS.ORG/BOOK

Preface

This work is part of a larger research program, Assumptions of Physics (https://assumptionsofphysics.org), that aims to identify a handful of physical principles from which the basic laws can be rigorously derived. The goal is to give physics (and science more in general) a renewed foundation that is mathematically precise, physically meaningful and philosophically consistent. Given the ambition and broad scope of the task, nothing would ever be written if one were to wait for the complete picture. Therefore this work contains only the parts of the project that are considered to be mature, and so it will be revised and expanded as progress is made. We give here a brief overview of the project, which can be useful to better understand the context of this work.

Overall goals of Assumptions of Physics

What do the basic laws of physics describe? Why is the state of a classical particle identified by position and momentum (i.e. a point on a symplectic manifold) while the state of a quantum system is identified by a wave function (i.e. a vector in a Hilbert space)? What assumptions are unique to each theory? What are, instead, the basic requirements that all physical theories must satisfy? Could we have had different laws? A lack of clear answers to these questions is, we believe, the biggest obstacle in the foundations of physics and prevents the resolution of outstanding problems in the field. **Our approach is to find a minimal set of physical assumptions that are necessary and sufficient to derive the different theories within a unified framework.** If we are able to do so, then we are guaranteed that all that the laws of physics say is encoded in those assumptions and we are able to answer those questions.

We found this approach to be very fruitful. It provides new insights into physics as a whole, the role of mathematics in physical theories and gives a more solid conceptual foundation to both. It becomes clear why some mathematical structures are pervasive in science and what exactly they are meant to represent, while others will never play a role. The downside is that we have to touch many subjects in math (logic, topological spaces, measure theory, group theory, vector spaces, differential geometry, symplectic geometry, statistics, probability theory, ...), physics (Hamiltonian mechanics, Lagrangian mechanics, thermodynamics, quantum mechanics, electromagnetism, ...) and science in general (computer science theory, information theory, system theory, ...). In other words, **the only way to properly achieve the goal is to rebuild everything from the ground up:** formal rigor, physical significance and conceptual integrity are not something that can be added at the end, but they must be present from the beginning.

The main takeaway for us is that the foundations of science are one: no real progress can be made on the foundations of one subject without making progress on the foundations of

others. **What is needed is a general theory of experimental science: a theory that studies physical theories.** This provides a standard framework that defines basic concepts and requirements (e.g. experimental verifiability, granularity of descriptions, processes and states) that serve as a common basis for all theories. Each theory, then, is recovered by studying how these common objects become specialized under different assumptions. This book aims to build over time, piece by piece, this framework.

While the topic is necessarily inter-disciplinary, **this is still first and foremost a scientific book**. The material should be accessible to the mathematician and philosopher, but understand that it needs to resonate first and foremost with the experimental physicist and the engineer. The mathematical definitions and derivations are there to make the science precise, but they are not the main focus. In fact, the book is designed so that the mathematical definitions and proofs, highlighted with a green side bar, can be skipped altogether without loss of the big picture and the important details. Along the same line, the foundational discussions are there to articulate more precisely what it means to do science, so they will not indulge in other questions which may be of interest to the philosopher but not to the scientist.

A living work

For the project to be successful, we need to depart from some of the norms of academic research and academic publishing. For example, one typically develops his research program as a series of articles published in a peer reviewed journal that caters to a specific community. These articles are then typically collected as is or merged into a book. This does not work for this project. As journals are specialized into sometimes very narrow fields, this would create a set of disjoint articles that cater to different audiences, with no guarantee that they can fit into a unified vision. For us, the overall picture, if and how the different perspectives combine, is the most important feature. **In this sense, the book comes first and the articles are derivative works.** We need to pool expertise and ideas from a wide range of disciplines and make sure that the result makes sense from all angles.

As the goal is broad, the framework needs to evolve as new issues are solved and old ones are better understood. If one part changes, we have to make sure that everything is updated to keep conceptual consistency. **This book, therefore, is an ongoing project.** It will continue to grow organically, adding and revising chapters. As is standard practice in open source/free software communities, we need to "release early, release often" to gather feedback. Each new version supersedes all prior ones and will be superseded by future ones. There is therefore no "definitive version" in the near future as we don't expect to "solve all of physics" in the near future. However, the framework will tend to converge as different parts become more settled.

The upshot is that one only needs to read the latest version of this work to be current. That is, one does not need to read a scattered set of papers, which require previous knowledge of a field, and follow how the ideas have changed. Just get the latest copy of the book, and if you do find areas you can help us expand or improve, let us know!

Project overview

Here we present a summary of the whole project and the status of each part as the layout will map to the structure of this work. We divide the work based on the two main techniques we use. The first, **reverse physics**, aims to identify the fundamental ideas and assumptions by reverse

engineering them from the current physical theories. The second, **physical mathematics**, aims to construct a rigorous mathematical framework from the ground up, based on the ideas and assumptions found by reverse physics.

Reverse physics

Reverse physics looks at the main physical theories, like classical mechanics, thermodynamics and quantum mechanics, to identify concepts that can be used to fully explain the common and different aspects of those theories. The core insight of reverse physics is that **when the math is derived from the right physical assumptions, the physical assumptions can be derived from the math**. That is, once one finds meaningful assumptions that can be shown to be equivalent to the physical laws, one can either start from the laws or the assumptions.

The standard of rigor in this part is necessarily more relaxed as we do not have a guarantee that sufficiently mature mathematical tools exist to carry out the argument in a precise fashion. For example, we have found that the idea of a unit system is linked in a fundamental way to the notion of state spaces, yet we lack a fully developed mathematical framework to model units and their dependency. The goal is to test the ideas conceptually, find those that are broad enough and necessary enough to then justify investing further time in a more rigorous approach.

The following are examples of the type of assumptions we have found to be good starting points to rederive the different theories.

Determinism and reversibility: "The system undergoes deterministic and reversible evolution." Mathematically, the physical properties of the system determine which category, in the mathematical sense, is used to describe the state space, and deterministic and reversible evolution will be an isomorphism in the category (i.e. a bijective map that preserves the physical properties of the system). Therefore the law of evolution is not just a bijective map, but is also a linear transformation, a differentiable map or an isometry depending on the context.

Infinitesimal reducibility: "Specifying the state of the whole system is equivalent to specifying the state of all its infinitesimal parts." For example, we can study the motion of a ball, but we can also mark a spot in red and study the motion of the mark. Knowing the evolution of the whole ball means knowing the evolution of any arbitrary spot and vice-versa. Mathematically, the state of the whole will be a distribution over the state space of the parts. It will need to be a distribution whose value is invariant under coordinate transformations. The state space of the infinitesimal parts, then, comes equipped with an invariant two-form upon which we can define such a distribution. The state space is therefore a symplectic manifold, that is, the states of the infinitesimal parts are described by pairs of conjugate variables, which recovers phase space. If the previous assumption holds, deterministic and reversible evolution is a symplectomorphism, that is, deterministic and reversible evolution follows classical Hamiltonian mechanics. Proper handling of the time variable will give us a relativistic version of the framework without extra assumptions.

Irreducibility: "Specifying the state of the whole system tells us nothing about its infinitesimal parts." For example, we can study the state of an electron by scattering photons off of it. But whenever a photon interacts with the electron, it interacts with the whole

electron. There is no way to mark a part of an electron and study it independently from the rest. Mathematically, the state of the electron will be a distribution that evolves deterministically where the motion of each infinitesimal part cannot be further described.

Kinematic equivalence: "Specifying the motion of the system is equivalent to specifying its state and evolution." This means that we will have to be able to re-express a distribution over kinematic variables (i.e. position and velocity) into a distribution over state variables (i.e. position and momentum) and vice-versa. Mathematically, the symplectic two-form will induce a symmetric tensor over the tangent space for position. This will give us a metric and will also allow us to reformulate the laws of motion according to Lagrangian mechanics. Because the transformation is linear, we are able to constrain the Hamiltonian to the one for massive particles under scalar and vector potential forces.

Physical mathematics

Physical mathematics aims to develop mathematical structures that are based on physical principles and assumptions, so that a perfect mapping exists between physical objects and their mathematical representation. The core insight of physical mathematics is that **when physical objects are mapped to the right mathematical objects, the physical requirements map to the mathematical definitions**. That is, the only way to have a perfect map between physical objects and their mathematical representation is if the mathematical axioms and definition can be justified by physical requirements.

As of now, we identified the following two core principles that serve as guidance to the development of the basic mathematical structures. As they describe requirements that any physical theory must satisfy, they are suitable to act as the foundation of a theory of scientific theories.

Principle of scientific objectivity. "Science is universal, non-contradictory and evidence based." This tells us that a scientific theory must be characterized by statements that are connected to experimental verification. Therefore, verifiable statements and their logic must provide a common foundation to all physical theories. Mathematically, these requirements are captured by topologies and σ-algebras over the space of the possible cases that can be identified experimentally. This part is very well developed both conceptually and mathematically.

Principle of scientific reproducibility. "Scientific laws describe relationships that can always be experimentally reproduced." This tells us that physical laws are relationships between inputs and outputs of repeatable procedures. Therefore, statistical ensembles provide a common foundation to define states and processes to all physical theories. Mathematically, a space of ensembles must be a topological space that allows convex combinations (i.e. statistical mixing) equipped with an entropy function that characterizes the variability of the elements within the ensemble. This part is still being developed, and will include open problems and conjectures.

Current plan and status

In this version we have added a physical mathematics chapter dedicated to ensemble spaces. For reverse physics, we plan to work on quantum mechanics. For physical mathematics, we plan to continue the work on ensemble spaces.

Changelog

2025/12/30: Ver 3.0 - Added chapter on ensemble spaces. Minor updates and corrections to old chapters.

2023/10/01: Ver 2.0 - Divided the work into two main parts: Reverse Physics and Physical Mathematics. Added chapter on reversing classical mechanics. Minor updates on the logic section.

2021/03/08: Ver 1.0 - Updated the first three chapters with minor changes: renamed tautology to certainty and contradiction to impossibility as they characterize better their role in the framework; made more formal justifications for the basic axioms and some of the basic definitions; causal relationships are now proved to be continuous instead of assumed to be continuous. Added Part II to include the results that are not yet fully formalized, to give a sense of the future scope of the work.

2019/07/07: Ver 0.3 - Reviewed first two chapters to clarify the idea of possible assignments and how contexts for function spaces are constructed.

2019/02/22: Ver 0.2 - Consolidated third chapter on properties, quantities and ordering.

2018/06/22: Ver 0.1 - Consolidated first two chapters that lay the foundation for the general theory.

Contents

Part I

Reverse Physics

Reverse physics is an approach to the foundations of physics that analyzes known theories to identify those physical principles and assumptions that can be taken as their conceptual foundation. It is based on the that **when the math is derived from the right physical assumptions, the physical assumptions can be derived from the math** This is the analogue of **reverse mathematics**, a program in mathematical logic that seeks to determine which axioms are required to prove mathematical theorems.

While some physical theories, such as Newtonian mechanics, thermodynamics and special relativity, are indeed founded on laws or principles, many theories, such as Hamiltonian and Lagrangian mechanics, quantum mechanics, and general relativity, are based on mathematical relationships that are simply postulated without a strict physical justification. Most modern theories are of the latter type, as theoretical physicists have increasingly focused on mathematical ideas rather than physical starting points. Therefore we do not know why the state of a quantum system can be represented by a ray in a Hilbert space, or what exact relationship is represented by the Einstein field equations. The goal of reverse physics, then, is to find suitable physical premises that can function as a more proper foundation for each theory. We stress that the premises have to be of a physical nature, such as "the system under study is isolated" or "the quantity is additive under system composition". That is, they must be principles or assumptions that express some physical idea, not some abstract mathematical notion.

Another issue in modern physics is that it is currently a patchwork of different theories, and one is trained to match patterns and examples to decide which problems (or what aspects of a larger problem) should be treated with a particular theory. The goal here is to fully understand what physical situation each mathematical structure can suitably describe, what connections can be made across different theories and what exactly are the true limits of applicability of the different physical theories. One of the main results of reverse physics is that there are some core ideas that are shared by all physical theories, and that, in fact, physical theories are not as disconnected from each other as generally thought.

While this work explores connections between different mathematical and physical disciplines, it would be impractical to provide even a concise introduction to all. Therefore when encountering concepts like information entropy, Minkowski space-time, function spaces, quantum observables, Lagrange multipliers, intensive quantities, and so on, we can only provide the definition with very limited context, as most of the space must be dedicated to the connections between these concepts. It is up to the reader to decide whether to invest additional time on other sources, or to simply make a mental note and proceed further.

By nature, reverse physics does not aim to be mathematically precise, but rather conceptually precise. While the goal is to create a dictionary between physical concepts and their mathematical representation, this mapping cannot be absolutely perfect for a simple reason: we have no guarantee that the current mathematical structures are the correct ones to capture the physical ideas. Conversely, we cannot give a precise mathematical characterization of the physical ideas if these are not conceptually clear. Therefore conceptual clarity has to come before formal precision. Yet, it is true that to reach full conceptual clarity the ideas have to be sufficiently refined to allow a formal definition. This task is the purview of **physical mathematics**, which aims to find physically meaningful mathematical structures starting from definitions and axioms that can be fully supported by physical requirements and assumptions. Yet, it would be premature to engage in such activity without having thoroughly tested the conceptual framework beforehand.

Chapter 1

Classical mechanics

The standard view in physics is that classical mechanics is perfectly understood. It has three different but equivalent formulations, the oldest of which, Newtonian mechanics, is based on three laws. Classical mechanics is the theory of point particles that follow those laws. Unfortunately, this view is incorrect.

We will see that the three formulations are not equivalent, in the sense that there are physical systems that are Newtonian but not Hamiltonian and vice-versa. There are also a number of questions that have been left unanswered, such as the precise nature of the Hamiltonian or the Lagrangian, and what exactly the principle of stationary action represents physically. While shedding light on these issues, we will also find that classical mechanics already contains elements that are typically associated with other theories, such as quantum mechanics/field theories (uncertainty principle, anti-particles), thermodynamics/statistical mechanics (thermodynamic and information entropy conservation) or special relativity (energy as the time component of a four-vector). In other words, the common understanding of classical mechanics is quite shallow, and its foundations are, in fact, not separate from the ones of classical statistical mechanics or special relativity.

What reverse physics shows is that the central assumption underneath classical mechanics is that of **infinitesimal reducibility (IR)**: a classical system can be thought of as made of parts, which in turn are made of parts and so on; studying the whole system is equivalent to studying all its infinitesimal parts. This assumption, together with the assumption of **independence of degrees of freedom (IND)**, is what gives us the structure of classical phase space with conjugate variables. The additional assumption of **determinism and reversibility (DR)**, the fact that the description of the system at one time is enough to predict its future or reconstruct its past, leads us to Hamiltonian mechanics. On the other hand, assuming **kinematic equivalence (KE)**, the idea that trajectories in space are enough to reconstruct the state of the system and vice-versa, leads to Newtonian mechanics. The combination of all above assumptions, instead, leads to Lagrangian mechanics and, in particular, to massive particles under (scalar and vector) potential forces.

As a guide to the chapter, here is the list of main points in the order in which they will be presented, one for each section.

1. Review of classical formulations
2. Lagrangian mechanics is Hamiltonian mechanics and KE
3. Kinematics, in general, is not enough to reconstruct dynamics

4. Hamiltonian mechanics (one DOF) is equivalent to DR
5. Hamiltonian mechanics (multiple DOFs) is equivalent to DR plus IND
6. Differential calculus and its generalization, differential topology, study infinitesimally additive quantities that depend on geometric shapes (i.e. lines, surfaces, volumes)
7. The principle of least action is a consequence of DR, IND and KE
8. Massive particles under potential forces are a consequence of DR, IND and KE
9. Special relativity is a consequence of DR, IND and KE
10. Phase space is the only structure that makes distributions, state counting and entropy frame invariant
11. Newtonian mechanics is a consequence of KE
12. Three dimensional spaces are the only spaces for which distributions over directions are frame invariant
13. Classical particle states as points in phase space are equivalent to IR

1.1 Formulations of classical mechanics

In this section we will briefly review the three main formulations of classical mechanics. Our task is not to present them in detail, but rather to provide a brief summary of the equations so that we can proceed with the comparison. In particular, given that different conventions are used across formulations, within the same formulation and among different contexts (e.g. relativity, symplectic geometry), we will want to make the notation homogeneous to allow easier comparisons.

Newtonian mechanics

For all formulations, the system is modeled as a collection of point particles, though we will mostly focus on the single particle case. For a Newtonian system, the state of the system at a particular time t is described by the position x^i and velocity v^i of all its constituents. Each particle has its mass m, not necessarily constant in time, and, for each particle, we define kinetic momentum as $\Pi^i = mv^i$.[1]

The evolution of our system is given by Newton's second law:[2]

$$F^i(x^j, v^k, t) = d_t \Pi^i. \tag{1.1}$$

Mathematically, if the forces F^i are locally Lipschitz[3] continuous, then the solution $x^i(t)$ is

[1] We will use the letter t for the time variable, x for position and v for velocity, which is a very common notation in Newtonian mechanics. However, we will keep using the same letters in Lagrangian mechanics as well, instead of q and $\dot{q}$, for consistency. Given that the distinction between kinetic and conjugate momentum is an important one, we will denote Π the former and p the latter. The Roman letters $i, j, k, \ldots$ will be used to span the spatial components (e.g. $i \in \{1, 2, 3\}$ for a particle in 3 dimensional space and $i \in \{1, 2, \ldots, 3n\}$ for n particles), while we will use the Greek letters $\alpha, \beta, \gamma, \ldots$ to span space-time components (e.g. $\alpha \in \{0, 1, 2, 3\}$ where the 0 value of the index is used for time). Unlike some texts, x^i do not represent Cartesian coordinates, and therefore they should be understood already as generalized coordinates.

[2] For derivatives, we will use the shorthand d_t for $\frac{d}{dt}$ and ∂_{x^i} for $\frac{\partial}{\partial x^i}$. For functions that depend on multiple arguments we use a free index to note that it depends on all elements; each argument will have a different index to highlight that there is no relationship between arguments.

[3] Lipschitz continuity means that the slope of the function is bounded. For example, $\sqrt{x}$ in the neighborhood of 0 is not Lipschitz continuous as it has a vertical asymptote at that point. One can construct examples (e.g. Norton's dome) where the forces are not locally Lipschitz continuous, and therefore the initial position and

unique. That is, given position and velocity at a given time, we can predict the position and velocity at future times. We will assume a Newtonian system has this property.

An important aspect of Newtonian mechanics is that the equations are not invariant under coordinate transformation. To distinguish between apparent forces (i.e. those dependent on the choice of frame) and the real ones, we assume the existence of inertial frames. In an inertial frame there are no apparent forces, and therefore a free system (i.e. no forces) with constant mass proceeds in a linear uniform motion, or stays still.[4]

Lagrangian mechanics

The state for a Lagrangian system is also given by position x^i and velocity v^i. The dynamics is specified by a single function $L(x^i, v^j, t)$ called the Lagrangian. For each spatial trajectory $x^i(t)$ we define the action as $\mathcal{A}[x^i(t)] = \int_{t_0}^{t_1} L(x^i(t), d_t x^i(t), t)dt$. The trajectory taken by the system is the one that makes the action stationary:

$$\delta\mathcal{A}[x^i(t)] = \delta \int_{t_0}^{t_1} L\left(x^i(t), d_t x^i(t), t\right) dt = 0 \tag{1.2}$$

The evolution can equivalently be specified by the Euler-Lagrange equations:

$$\partial_{x^i} L = d_t \partial_{v^i} L. \tag{1.3}$$

Note that not all Lagrangians lead to a unique solution. For example, $L = 0$ will give the same action for all trajectories and therefore, strictly speaking, all trajectories are possible. The stationary action leads to a unique solution if and only if the Lagrangian is hyperregular, which means the Hessian matrix $\partial_{v^i}\partial_{v^j} L$ is invertible. Like in the Newtonian case, we will assume Lagrangian systems satisfy this property.

Unlike Newton's second law, both the Lagrangian and the Euler-Lagrange equations are invariant under coordinate transformations. This means that Lagrangian mechanics is particularly suited to study the symmetries of the system.

Hamiltonian mechanics

In Hamiltonian mechanics, the state of the system is given by position q^i and conjugate momentum p_i. The dynamics is specified by a single function $H(q^i, p_j, t)$ called the Hamiltonian.[5] The evolution is given by Hamilton's equations:

$$\begin{aligned} d_t q^i &= \partial_{p_i} H \\ d_t p_i &= -\partial_{q^i} H \end{aligned} \tag{1.4}$$

velocity do not yield a unique solution (i.e. in Norton's dome, the body can stay on the top of the dome indefinitely, or it can fall down after an arbitrary amount of time). In this case, something else, outside the system, will necessarily determine what is the motion of the system, and therefore it is not true that the force and the state of the system fully determine the dynamics of the system.

[4] Recall that linear motion simply means that it describes a line in space, while uniform motion means that the speed is constant. Therefore we can have linear non-uniform motion (e.g. an object accelerated along the same direction) or a non-linear uniform motion (e.g. an object going around in a circle at constant speed).

[5] We use a different symbol for position in Hamiltonian mechanics because, while it is true that $q^i = x^i$, it is also true that $\partial_{q^i} \neq \partial_{x^i}$: the first derivative is taken at constant conjugate momentum while the second is taken at constant velocity. This creates absolute confusion when mixing and comparing Lagrangian and Hamiltonian concepts, which our notation avoids completely.

We will again want these equations to yield a unique solution, which means the Hamiltonian must be at least differentiable, and the derivatives must at least be Lipschitz continuous.

Hamilton's equations are invariant as well. The Hamiltonian itself is a scalar function which is often considered (mistakenly as we'll see later) invariant. This formulation is the most suitable for statistical mechanics as volumes of phase space correctly count the number of possible configurations.

1.2 Inequivalence of formulations

It is often stated in physics books that all three formulations of classical mechanics are equivalent. We will look at this claim in detail, and conclude that this is not the case: there are systems that can be described by one formulation and not another. More precisely, the set of Lagrangian systems is exactly the intersection of Newtonian and Hamiltonian systems.

Testing equivalence

We will consider two formalisms equivalent if they can be applied to exactly the same systems. That is, Newtonian and Lagrangian mechanics are equivalent if any system that can be described using Newtonian mechanics can also be described by Lagrangian mechanics and vice-versa. In general, in physics great emphasis is put on systems that can indeed be studied by all three, leaving the impression that this is always doable.[6] However, just with a cursory glance, we realize that this can't possibly be the case.

The dynamics of a Newtonian system, in fact, is specified by three independently chosen functions of position and velocity, the forces applied to each degree of freedom (DOF). On the other hand, the dynamics of Lagrangian and Hamiltonian systems is specified by a single function of position and velocity/momentum, the Lagrangian/Hamiltonian. Intuitively, there are more choices in the dynamics for Newtonian systems than for Lagrangian and Hamiltonian.

Now, the reality is a bit trickier because the mathematical expression of the forces is not enough to fully characterize the physical system. We need to know in which frame we are, what coordinates are being used and the mass of the system, which is potentially a function of time. On the Lagrangian side, note that the Euler-Lagrange equations are homogeneous in L. This means that multiplying L by a constant leads to the same solutions, meaning that the same system can be described by more than one Lagrangian. The converse is also true: if one system is half as massive and is subjected to a force half as intense, the resulting Lagrangian is also simply rescaled by a constant factor. Therefore the map between Lagrangians and Lagrangian systems is not one-to-one: it is many-to-many. This is why we should never look simply at mathematical structures if we want to fully understand the physics they describe.

Regardless, our task is at the moment much simpler: we only need to show that there are Newtonian systems not expressible by Lagrangian or Hamiltonian mechanics. We can therefore limit ourselves to systems with a specific constant mass m in an inertial frame and write $a^i = F^i(x^j, v^k, t)/m$. Given that the force is arbitrary, the acceleration can be an arbitrary function of position, velocity and time. Similarly, we can write the acceleration of a

[6] If one asks the average physicist whether Newtonian and Hamiltonian mechanics are equivalent, the answer most of the time will be enthusiastically positive. If one then asks for the Hamiltonian for a damped harmonic oscillator, the typical reaction is annoyance due to the nonsensical question (damped harmonic oscillators do not conserve energy), followed by a realization and partial retraction of the previous claim. The moral of the story is to never take these claims at face value.

Lagrangian system as $a^i = F^i[L]/m$. That is, the acceleration is going to be some functional of the Lagrangian. Given the Euler-Lagrange equations 1.3, the map between the Lagrangian and the acceleration must be continuous in both directions: for a small variation of the Lagrangian we must have a small variation of the equations of motion and therefore of the acceleration, and for a small variation of the equations of motion we must have a small variation of the Lagrangian. But a continuous surjective map from the space of a single function (i.e. the Lagrangian) to the space of multiple functions (i.e. those that specify the acceleration in terms of position and velocity) does not exist,[7] and therefore there must be at least one Newtonian system with constant mass expressed in an inertial frame that is not describable using Lagrangian mechanics. The same argument applies for Hamiltonian mechanics, since the dynamics in this case is also described by a single function in the same number of arguments. We therefore reach the following conclusion:

Insight 1.5. *Not all Newtonian systems are Lagrangian and/or Hamiltonian.*

Newtonian vs Lagrangian/Hamiltonian

We now want to understand whether all Lagrangian systems are Newtonian. Given what we discussed, we cannot expect to reconstruct the mass and force uniquely from the expression of the Lagrangian. We consider the mass and the frame fixed by the problem, together with the Lagrangian, and therefore we must only see whether we can indeed find a unique expression for the acceleration. From the Euler-Lagrange equations 1.3 we can write

$$\begin{aligned} \partial_{x^i} L &= d_t \partial_{v^i} L = \partial_{x^j} \partial_{v^i} L \, d_t x^j + \partial_{v^k} \partial_{v^i} L \, d_t v^k + \partial_t \partial_{v^i} L \, d_t t \\ &= \partial_{x^j} \partial_{v^i} L \, v^j + \partial_{v^k} \partial_{v^i} L \, a^k + \partial_t \partial_{v^i} L \\ \partial_{v^k} \partial_{v^i} L \, a^k &= \partial_{x^i} L - \partial_{x^j} \partial_{v^i} L \, v^j - \partial_t \partial_{v^i} L. \end{aligned} \tag{1.6}$$

To be able to write the acceleration explicitly, we must be able to invert the Hessian matrix $\partial_{v^k} \partial_{v^i} L$. As we noted before, this is exactly the condition for which the principle of stationary action leads to a unique solution, and we can better understand why. If it is not invertible at a point, the determinant is zero and therefore one eigenvalue is zero. The corresponding eigenvector corresponds to a direction for which the equation tells us nothing, and therefore a variation of the acceleration in that direction will not change the action. This is why the invertibility of the Hessian is required in order to obtain unique solutions.

What we find, then, is that for any Lagrangian system, which we assume to have a unique solution, we can explicitly write the acceleration as a function of position, velocity and time. Therefore

Insight 1.7. *All Lagrangian systems are Newtonian.*

Now we turn our attention to Hamiltonian mechanics and, similarly, we ask whether we can express the acceleration as a function of the state. We have

$$\begin{aligned} a^i &= d_t v^i = d_t d_t q^i = d_t \partial_{p_i} H = \partial_{q^j} \partial_{p_i} H \, d_t q^j + \partial_{p_k} \partial_{p_i} H \, d_t p_k + \partial_t \partial_{p_i} H \, d_t t \\ &= \partial_{q^j} \partial_{p_i} H \partial_{p_j} H - \partial_{p_k} \partial_{p_i} H \partial_{q^k} H + \partial_t \partial_{p_i} H. \end{aligned} \tag{1.8}$$

[7] Mathematically, the space of continuous functions $C(\mathbb{R}, \mathbb{R})$ and $C(\mathbb{R}^n, \mathbb{R})$ are not homeomorphic. Intuitively, the underlying reason is the same as to why a map from a volume to a line can't be continuous: in a volume you have infinitely many directions you can move away from a point, while on a line you only have two.

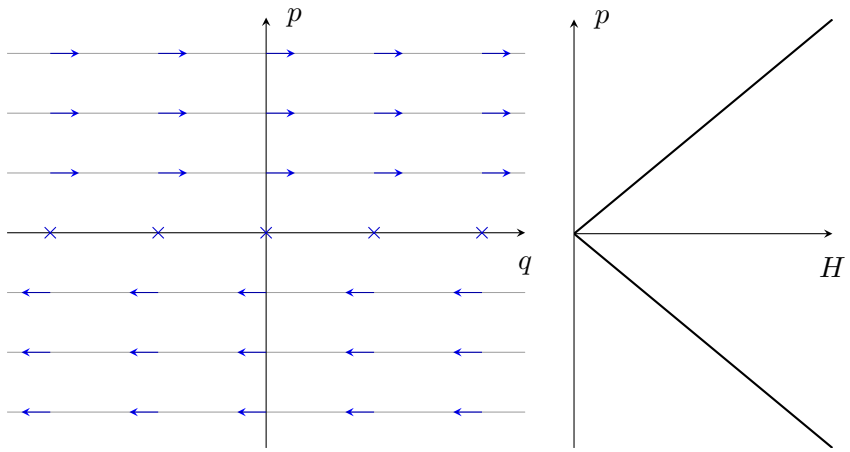


Figure 1.1: On the left, the phase-space diagram for a photon treated as a point particle. The Hamiltonian $H = c|p|$, on the right, is proportional to the modulus of p. Since H is not differentiable when $p = 0$, those states are excluded, consistent with the physics. The displacement field has only a q component, which is $+c$ above the horizontal axis and $-c$ below the horizontal axis.

This tells us that we can always write an explicit function for the acceleration. However, this is not enough. States in Newtonian mechanics are in terms of position and velocity, not position and momentum. For a Hamiltonian system to be equivalent to a Newtonian system we need to be able to write the momentum as a function of position and velocity and vice versa. Note that Hamilton's equations already give a way to express the velocity in terms of position and momentum. We just need that expression to be invertible, which means the Jacobian must be invertible. We must have:

$$\left|\partial_{p_i} v^j\right| = \left|\partial_{p_i}\partial_{p_j} H\right| \neq 0. \tag{1.9}$$

To be able to express momentum as a function of position and velocity, then, we need the Hessian of the Hamiltonian to be invertible (i.e. to have non-zero determinant).

Note that we had no such requirement for the Hamiltonian itself. For example, $H = 0$ leads to equations $d_t q^i = 0$ and $d_t p_i = 0$, which have unique solutions: both position $q^i(t) = k_{q^i}$ and momentum $p_i(t) = k_{p_i}$ are constants of motion. The Hessian, being the zero matrix, is not invertible, and in fact we cannot write momentum as a function of position and velocity: velocity $d_t q^i$ is always zero in all cases while conjugate momentum can be any value k_{p_i}. Though this case may not be physically interesting, it is a perfectly valid Hamiltonian system and shows that we should always check the trivial mathematical case. However, let us go through a more physically meaningful case.

Photon as a particle. If we want to treat the photon as a classical particle, we can write the Hamiltonian by expressing the energy as a function of momentum

$$H = \hbar|\omega| = c\hbar|k_i| = c|p_i|. \tag{1.10}$$

If we apply Hamilton's equations, we have

$$\begin{aligned} d_t q^i &= c\frac{p_i}{|p_i|} \\ d_t p_i &= 0. \end{aligned} \tag{1.11}$$

That is, the norm of the velocity is always c, the momentum decides its direction, and the momentum itself does not change in time, as shown in fig. 1.1. This is indeed the motion of a free photon. One can confirm, through tedious calculation, that the determinant of the Hessian is indeed zero, yet it is easier and more physically instructive to see that we cannot reconstruct the momentum from the velocity. Relativistically, all photons travel along the geodesics at the same speed, therefore two photons that differ only by the magnitude of the momentum will travel the same path.

Hamiltonian systems that are also Newtonian, then, need to satisfy this extra condition, so let us give it a name.

Assumption KE (Kinematic Equivalence)**.** *The kinematics of the system is sufficient to reconstruct its dynamics and vice-versa. That is, specifying the motion of the system is equivalent to specifying its state and evolution.*

By kinematics we mean the motion in space and time and by dynamics we mean the state and its time evolution in phase space. We will need to analyze the difference between the two more in detail, but we should first finish our comparison between the different formulations.

Summing up, we find that

Insight 1.12. *Not all Hamiltonian systems are Newtonian: only those for which KE is valid.*

Lagrangian vs Hamiltonian

We now need to compare Lagrangian and Hamiltonian systems. The task is a lot easier because we already have a precise way to connect the two. If we are given a Lagrangian L, we define the conjugate momentum $p_i = \partial_{v^i} L$ and the Hamiltonian $H = p_i v^i - L$. If we are given a Hamiltonian H, we can define a Lagrangian $L = p_i v^i - H$ and a velocity $v^i = d_t q^i = \partial_{p_i} H$. However, this is a bit misleading: the above relationships connect the values of the functions for each state s. That is, $L(s) = p_i(s)v^i(s) - H(s)$. Both the Lagrangian and the Hamiltonian are functions of specific variables, so we have to make sure we can express them in the appropriate variables.

Going from a Hamiltonian to a Lagrangian, it again means that we can write momentum as a function of position and velocity, and therefore assumption KE must hold. This makes sense: if all Lagrangian systems are Newtonian, and KE was required for a Hamiltonian system to be Newtonian, then it is also required for a Hamiltonian system to be Lagrangian. But the connection is stronger: KE is the *only* additional assumption we need to be able to write a Lagrangian given a Hamiltonian.

Going from a Lagrangian to a Hamiltonian, it means that we can write velocity as a function of position and momentum. Note that since we define conjugate momentum as the derivative of the Lagrangian, we can already express momentum as a function of position and velocity, which means we are simply asking that expression to be invertible. This is, again, assumption KE, just in the opposite direction. We must have

$$0 \neq |\partial_{v^i} p_j| = |\partial_{v^i}\partial_{v^j} L|\,. \tag{1.13}$$

This means that assumption KE is exactly the invertibility of the Hessian, the condition for unique solution of the Lagrangian. All Lagrangian systems that admit unique solutions, then, satisfy assumption KE. In fact, we can see that the Hessian determinants are related

$$|\partial_{v^i}\partial_{v^j}L| = |\partial_{v^i}p_j| = \left|\partial_{p_i}v^j\right|^{-1} = \left|\partial_{p_i}\partial_{p_j}H\right|^{-1}. \tag{1.14}$$

This means that every Lagrangian admits a Hamiltonian, but not every Hamiltonian admits a Lagrangian. Only the Hamiltonian systems for which KE is valid will also be Lagrangian systems, with a guaranteed unique solution given that KE is exactly the assumption needed for that as well. Therefore we conclude that

Insight 1.15. *Lagrangian systems are exactly those Hamiltonian systems for which KE is valid.*

Relationship between formulations

The relationship between the different formulations, then, can be summarized with the Venn diagram in fig. 1.2.

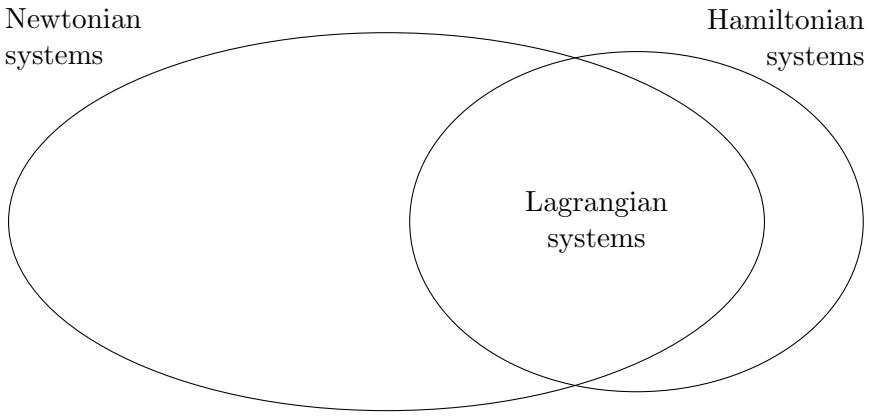


Figure 1.2: Not all Hamiltonian systems are Newtonian and not all Newtonian systems are Hamiltonian. All Lagrangian systems are both Newtonian and Hamiltonian.

We have found that KE is a constitutive assumption of Lagrangian mechanics, and that it clearly marks which Hamiltonian systems are Newtonian/Lagrangian. By constitutive assumption we mean an assumption that must be taken, either explicitly or implicitly, for a theory to be valid. But what makes a system Hamiltonian and what makes a system Newtonian? Can we find a full set of constitutive assumptions for classical mechanics?

1.3 Kinematics vs dynamics

We have seen the importance of the connection between kinematics and dynamics. In this section we will explore this link more deeply and come to the following conclusion: the kinematics of a system is not enough to reconstruct its dynamics.

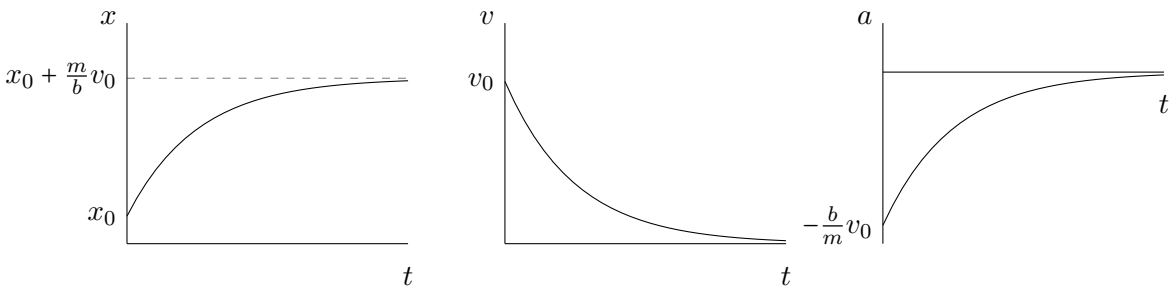


Figure 1.3: Evolution in time of position, velocity and acceleration for $ma = -bv$. Both acceleration and velocity will tend to zero as time increases. The position will tend to an equilibrium given by initial position and initial velocity.

Particle under linear drag

Let us first review exactly what the kinematics and dynamics are. Given a system, its kinematics is the description of its motion in space and time. Position, velocity, and acceleration are kinematic variables because they describe the motion. Kinematics is what Galileo studied and started to give a rigorous account of. The dynamics, instead, describes the cause of such motion. Force, mass, momentum, energy are dynamic quantities as they are used to describe why a body moves in a particular way. Dynamics is what Newton introduced and his second law, expressed as $F = ma$, clearly shows the link.

The link between the two concepts seems important given the constitutive role of KE in Lagrangian mechanics. Moreover, while both Newtonian and Hamiltonian mechanics are dynamical theories, in the sense that quantities like force and momentum are intrinsic parts of the respective theories, Lagrangian mechanics seems to be a purely kinematic theory, as it is described only by kinematic variables like position and velocity. Therefore it seems useful to characterize the kinematics-dynamics link as much as possible. Let's analyze a concrete example.

Suppose we are given the following equation:

$$ma = -bv. \tag{1.16}$$

The equation is in terms of kinematic variables and, given initial conditions x_0 and v_0, it admits a unique solution, a unique trajectory. The solution, plotted in fig. 1.3, is

$$\begin{aligned} x(t) &= x_0 + v_0 \frac{m}{b}\left(1 - e^{-\frac{b}{m}t}\right) \\ v(t) &= v_0 e^{-\frac{b}{m}t} \\ a(t) &= -v_0 \frac{b}{m} e^{-\frac{b}{m}t} \end{aligned} \tag{1.17}$$

Can we reconstruct the forces acting on this system?

The obvious answer seems to be that the constant m represents the mass of the system and $F = -bv$ the force. This is the case of a particle under linear drag: the system is subjected to a frictional force that is proportional and opposite to the velocity. If we set the Lagrangian

$$L = \frac{1}{2} m v^2 e^{\frac{b}{m}t}. \tag{1.18}$$

and apply the Euler-Lagrange equation 1.3 we have

$$\begin{aligned} \partial_x L = 0 = d_t \partial_v L = d_t \left(mve^{\frac{b}{m}t} \right) = mae^{\frac{b}{m}t} + \frac{b}{m} mve^{\frac{b}{m}t} = e^{\frac{b}{m}t}(ma + bv) \\ ma = -bv. \end{aligned} \tag{1.19}$$

Therefore we have a Lagrangian for the system. We can also find a Hamiltonian

$$\begin{aligned} p &= \partial_v L = mve^{\frac{b}{m}t} \\ v &= \frac{p}{m} e^{-\frac{b}{m}t} \\ H &= pv - L = p\frac{p}{m}e^{-\frac{b}{m}t} - \frac{1}{2}m\left(\frac{p}{m}e^{-\frac{b}{m}t}\right)^2 e^{\frac{b}{m}t} = \frac{p^2}{m}e^{-\frac{b}{m}t} - \frac{1}{2}\frac{p^2}{m}e^{-\frac{b}{m}t} \\ &= \frac{1}{2}\frac{p^2}{m}e^{-\frac{b}{m}t} \end{aligned} \tag{1.20}$$

and apply Hamilton's equations 1.4

$$\begin{aligned} d_t q &= \partial_p H = \frac{p}{m} e^{-\frac{b}{m}t} \\ d_t p &= -\partial_q H = 0. \end{aligned} \tag{1.21}$$

The second equation tells us momentum is constant p_0. Substituting the constant in the first equation, we have the velocity as a function of time, which we can integrate. We have

$$\begin{aligned} q(t) &= q_0 + \frac{p_0}{b}\left(1 - e^{-\frac{b}{m}t}\right) \\ p(t) &= p_0. \end{aligned} \tag{1.22}$$

The kinematics works perfectly, but the dynamics seems off, as shown in fig. 1.4. First of all, based on the physics, one would expect the momentum to be decreasing in time

$$p(t) = mv(t) = mv_0 e^{-\frac{b}{m}t}. \tag{1.23}$$

However, conjugate momentum is a constant of motion. For the energy, we would expect the Hamiltonian to match the kinetic energy

$$E(t) = \frac{1}{2}mv^2(t) = \frac{1}{2}mv_0^2 e^{-2\frac{b}{m}t} \tag{1.24}$$

but if we express the Hamiltonian in terms of velocity we have

$$H(t) = \frac{1}{2}\frac{p^2}{m}e^{-\frac{b}{m}t} = \frac{1}{2}\frac{1}{m}\left(mv(t)e^{\frac{b}{m}t}\right)^2 e^{-\frac{b}{m}t} = \frac{1}{2}mv^2(t)e^{\frac{b}{m}t} = \frac{1}{2}mv_0^2 e^{-\frac{b}{m}t}. \tag{1.25}$$

That is, the energy decreases more slowly than it should. This is not good.

Now, it is true that conjugate momentum is not the same as kinetic momentum. But the difference, as we will see much more clearly later, is caused by non-inertial non-Cartesian coordinate systems and/or the presence of vector potential forces.[8] We are not at all in that case. Also, note that at time $t = 0$ the momentum and the energy do match our expectation, but not after. Therefore imagine a situation where friction is non-negligible only in a particular region. We would expect $p = mv$ to be valid before it enters, but not when it comes out. But wouldn't it come out in another region where we would expect $p = mv$ to work? This is strange. How should we proceed?

[8] The relationship is $p_i = mg_{ij}v^j + \mathfrak{q}A_i$. This reduces to $p_i = mv^i$ if and only if we are in an inertial frame with Cartesian coordinates (i.e. $g_{ij} = \delta_{ij}$) and no forces $A_i = 0$

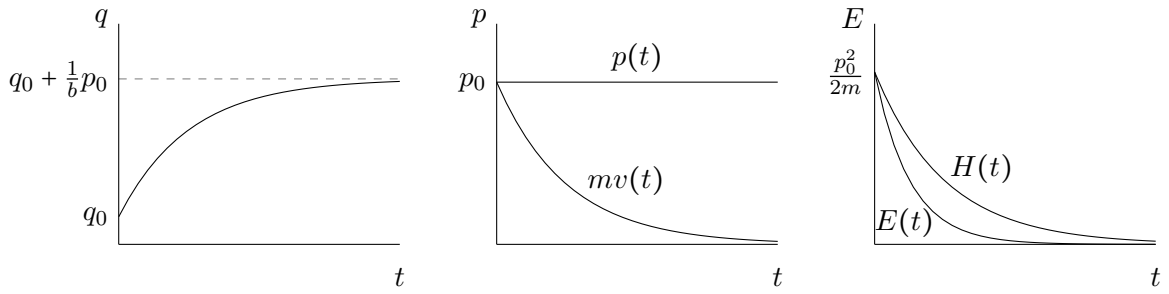


Figure 1.4: Trying to interpret $L = \frac{1}{2}mv^2e^{\frac{b}{m}t}$ and $H = \frac{1}{2}\frac{p^2}{m}e^{-\frac{b}{m}t}$ as respectively the Lagrangian and Hamiltonian of a particle under linear drag. While evolution of the position matches, note how the conjugate momentum is constant while the kinetic momentum decreases. Also, the Hamiltonian and the energy do not decrease at the same rate.

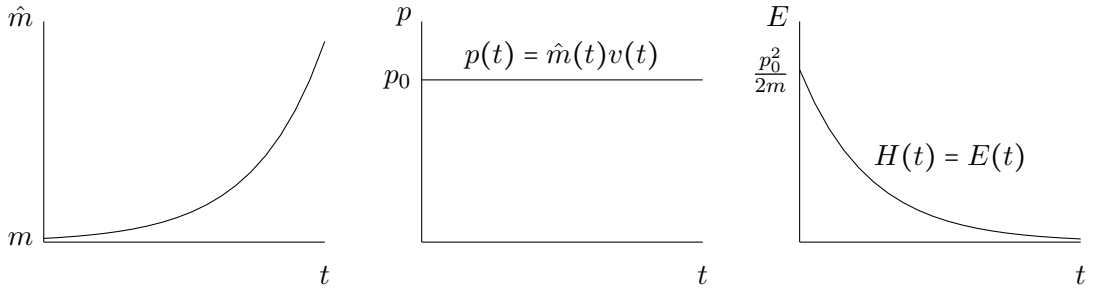


Figure 1.5: Showing how $L = \frac{1}{2}mv^2e^{\frac{b}{m}t}$ and $H = \frac{1}{2}\frac{p^2}{m}e^{-\frac{b}{m}t}$ can be interpreted as the Lagrangian and Hamiltonian of a variable mass system. The mass is increasing exponentially in time, while both conjugate and kinetic momentum remain constant. This means the velocity will need to decrease. The energy decreases at the same rate as the Hamiltonian.

Variable mass system

As it is typical in reverse physics, we will assume that things work in a reasonable way and that we simply have the wrong connection between physics and math. Recall that we started just with an equation, and we then interpreted m to be the mass of the system. Let's just assume that m is a constant with units of mass and define the actual mass of the system as the ratio between conjugate momentum and velocity. Looking back at 1.20, as shown in fig. 1.5, we have

$$\begin{aligned}
\hat{m}(t) &= p(t)/v(t) = me^{\frac{b}{m}t} \\
p(t) &= mv(t)e^{\frac{b}{m}t} = \hat{m}(t)v(t) \\
H(t) &= \frac{1}{2}\frac{p^2(t)}{m}e^{-\frac{b}{m}t} = \frac{1}{2}\frac{p^2(t)}{\hat{m}} = \frac{1}{2}\hat{m}(t)v^2(t) = E(t)
\end{aligned} \tag{1.26}$$

Now everything actually works perfectly: the relationship between velocity and conjugate momentum is respected, the Hamiltonian matches the kinetic energy. We just have a variable mass system. How and why does this work exactly?

Let us expand Newton's second law for a variable mass system.[9] We have:

$$\begin{aligned} F^i &= d_t(\hat{m}v^i) = d_t\hat{m}\, v^i + \hat{m}a^i \\ \hat{m}a^i &= F^i - d_t\hat{m}v^i \end{aligned} \tag{1.27}$$

In particular, for our one dimensional case, let us set $F = 0$ and substitute $\hat{m}$

$$\begin{aligned} me^{\frac{b}{m}t}a = 0 - d_t me^{\frac{b}{m}t}v &= -\frac{b}{m}me^{\frac{b}{m}t}v \\ ma &= -bv. \end{aligned} \tag{1.28}$$

Therefore the same equation, the same kinematics, applies to a variable mass system that increases the mass over time. You can imagine, for example, a body that is absorbing mass from all directions, so that the balance of forces on the body is zero. The body, then, is not slowing down because of friction. It is slowing down because momentum is conserved, and if the mass is increasing, the velocity must be decreasing at the same rate. The energy, on the other hand, will decrease because the square of the velocity will decrease faster than the mass increases.

In Newtonian mechanics, we can readily distinguish these two cases because we have to be explicit about forces and masses. In Hamiltonian mechanics things are a bit more difficult because, as we will see later more precisely, conjugate momentum is not exactly kinetic momentum and the Hamiltonian is not exactly energy. Yet, conjugate momentum and the Hamiltonian are not kinematic quantities, they are dynamic quantities and therefore we can see that these would be different in different cases. In Lagrangian mechanics this is even more difficult to see because it looks like a purely kinematic theory, while it is not: the Lagrangian itself is not a purely kinematic entity. As we saw, Lagrangian mechanics implicitly assumes KE, which is a condition on the dynamics as well, and the Lagrangian itself is used to reconstruct conjugate momentum and the Hamiltonian. Moreover, if Lagrangian mechanics were a purely kinematic theory, and told us nothing about forces, energy or momentum, it would not be a complete formulation of classical mechanics.

So we have seen that the same kinematic equation can describe a constant mass dissipative system or a variable mass system. Is that it? Not quite. Recall that we mentioned that kinetic and conjugate momentum will differ in non-inertial frames. Note that we implicitly assumed that x and t represented the variables for an inertial observer, in the same way that we originally assumed m was the mass of the system. Could the same equation, then, be describing yet another system but in a non-inertial frame?

Non-inertial motion

Let's compare the motion of a particle traveling at constant velocity in an inertial frame, using t as the time variable, and the motion of a particle decelerating exponentially, using $\hat{t}$ as the time variable

$$\begin{aligned} x(t) &= x_0 + v_0 t \\ x(\hat{t}) &= x_0 + v_0\frac{m}{b}\left(1 - e^{-\frac{b}{m}\hat{t}}\right). \end{aligned} \tag{1.29}$$

[9] Note that, in general, the variable mass system should take into account the momentum gained or lost by the system when the mass is acquired or ejected. In our case, we are assuming that no momentum is lost, which means that either the mass is acquired/ejected uniformly from all directions or it is just an apparent change that depends on the change of coordinates.

Note the striking similarity: we can simply set

$$t = \frac{m}{b}\left(1 - e^{-\frac{b}{m}\hat{t}}\right) \tag{1.30}$$

which clearly takes us to a non-inertial frame since uniform motion is no longer uniform in the new frame.

Let's study how Newton's second law changes if we make a change of time variable while keeping the position variables unchanged

$$\begin{aligned} &\hat{t} = \hat{t}(t) \\ &F^i = d_t(m\, v^i) = d_t(m\, d_t x^i) = d_t\hat{t}\, d_{\hat{t}}(m\, d_t\hat{t}\, d_{\hat{t}} x^i). \end{aligned} \tag{1.31}$$

If we set

$$\hat{m} = m\, d_t\hat{t} \tag{1.32}$$

we can express the previous equation in the following form

$$F^i = d_t\hat{t}\, d_{\hat{t}}(\hat{m}\, d_{\hat{t}} x^i) = d_t\hat{t}\, d_{\hat{t}}(\hat{m}\, \hat{v}^i) = d_t\hat{t}\hat{F}^i. \tag{1.33}$$

This tells us that the second observer will see an effective mass rescaled exactly by the ratio between the time variables. Note that this is exactly what happens in special relativity: the clock for a boosted observer is dilated by a factor of γ which is exactly the factor used in the relativistic mass.[10] If t is the time variable for an inertial frame and $t(\hat{t})$ is a non-linear function, the resulting frame will be non-inertial and the observer will see an effective variable mass system.

If we look at our problem this way, the rescaling of the mass, then, is not due to a truly variable mass, but a variable effective mass due to the slowing down of the clock. The body slows down because the non-inertial time is slowing down and the body appears to stop because the clock becomes infinitely slow. While this might sound like a contrived case,[11] these are exactly the type of situations a fully relativistic theory (i.e. one that works for all definitions of time and space variables) needs to take into account.

We can verify that this gives us the correct effective mass

$$\begin{aligned} &d_{\hat{t}} t = d_{\hat{t}}\left(\frac{m}{b}(1 - e^{-\frac{b}{m}\hat{t}})\right) = \frac{m}{b} d_{\hat{t}}(1 - e^{-\frac{b}{m}\hat{t}}) = -\frac{m}{b} d_{\hat{t}} e^{-\frac{b}{m}\hat{t}} = +\frac{m}{b}\frac{b}{m} e^{-\frac{b}{m}\hat{t}} = e^{-\frac{b}{m}\hat{t}} \\ &\hat{m} = m d_t\hat{t} = m(d_{\hat{t}} t)^{-1} = m e^{\frac{b}{m}\hat{t}}. \end{aligned} \tag{1.34}$$

And we can verify that we get the same equation by plugging in the time transformation in Newton's second law with a zero force

$$\begin{aligned} 0 &= d_t(m\, v) = d_t(m\, d_t x) = d_t\hat{t}\, d_{\hat{t}}(m\, d_t\hat{t}\, d_{\hat{t}} x) \\ &= e^{\frac{b}{m}\hat{t}}\, d_{\hat{t}}(m e^{\frac{b}{m}\hat{t}}\, \hat{v}) = e^{\frac{b}{m}\hat{t}}\left(m\frac{b}{m} e^{\frac{b}{m}\hat{t}}\hat{v} + m e^{\frac{b}{m}\hat{t}}\hat{a}\right) = e^{2\frac{b}{m}\hat{t}}\left(b\hat{v} + m\hat{a}\right) \\ m\hat{a} &= -b\hat{v}. \end{aligned} \tag{1.35}$$

[10] It may be surprising to see a proto-relativistic effect showing up given that no assumption on space-time has been made. As we will see, these types of connections between different theories come up often in reverse physics.

[11] On the surface, it sounds similar to what happens in general relativity with a black-hole. An observer that sees someone falling into a black hole will see him gradually slowing down as he approaches the event horizon and asymptotically stop there. The observer falling inside the black hole, instead, will perceive his time flowing uniformly and nothing special will happen as the event horizon is crossed.

Note that the expressions for momentum and energy will match the previous case because the system in the non-inertial frame looks like a variable mass system.

The relationship between kinematics and dynamics

This last case highlights a more subtle issue. In the two previous cases we were in the same inertial frame, we saw the same trajectory, the same kinematics, but we couldn't tell whether we were looking at a fixed mass system under linear drag or a variable mass system: we couldn't tell the dynamics. Now, we have the same system, a constant mass particle under no forces, described in two different frames, one inertial and one not. The motion of the system will naturally have different representations in the different frames, but this does not mean the motion or the causes of motion are different: it's the same object. Therefore we have the same motion even though we have different expressions for the trajectory. The expression $x(t)$, then, is not enough to define the kinematics if we do not know exactly what x and t represent physically, if the frame is not given.

While typically one proceeds by defining the frame first and then the dynamics (i.e. the forces acting on the system), here we have followed a different approach: we first defined the dynamics (i.e. constant mass system under no forces) and then found the frame that matched the given kinematics (i.e. the trajectory or the relationship between velocity and acceleration). Given that Lagrangian and Hamiltonian mechanics are frame invariant, an intrinsic characterization of the system itself is exactly what we should be looking for. Saying, for example, that a system is subjected to no forces or to a linear drag is not frame invariant because forces are not frame invariant.

It is clear that the type of apparent variable mass due to non-inertial frames is unavoidable if we want to have a consistent theory with invariant laws. Therefore both Lagrangian and Hamiltonian mechanics must include these cases. However, it is not exactly clear what to do for true variable mass systems. From a cursory look, it would seem that everything is fine and there is no harm in including them. Yet again, from a cursory look we seemed to have a Lagrangian for a particle under linear drag. As we will see later, there are implicit connections between Lagrangian/Hamiltonian mechanics on one side and thermodynamics, statistical mechanics and special relativity on the other. Given that it is not clear to us whether these connections hold or not,[12] we will concentrate on the constant mass case from now on.

Let's recap what we learned. The biggest point is that we can't simply look at the kinematics and understand the causes of motion. The different formulations have different ways to relate the dynamics and the kinematics. Newtonian mechanics is the most clear about the dynamics as it makes us clearly spell out what is going on. This, however, comes at a cost: the equations are not covariant, meaning they have a different expression in different frames. The second law, in fact, is valid only for inertial frames with Cartesian coordinates. It is only in these frames, in fact, that a body will proceed in uniform motion if no forces are applied to it. If we are in polar coordinates, for example, the trajectory expressed in radius r and angle θ will not be linear. Even the notion of force is, if one looks closely, a bit ambiguous. In principle, we want to write both the second law $F = ma$ and the expression for work $dW = Fdx$. If dW is invariant under change of position variables, the force should be a covector and therefore $dW = F_i dx^i$. But since the acceleration a will change like a vector, we also have $F^i = ma^i$.

[12] For example, areas of phase space are connected to entropy. Does this connection hold with a variable mass system?

The notion of force in the second law and in the infinitesimal work are slightly different, and they coincide only if we are in an inertial frame and Cartesian coordinates.

On the other side, Hamiltonian and Lagrangian mechanics are coordinate independent: the laws remain the same if we change position variables. This makes them more useful in many contexts. Lagrangian mechanics is more useful when trying to study the symmetries of the system. Hamiltonian mechanics is more useful for statistical mechanics and to better separate degrees of freedom. However, this comes at a price. Hamiltonian and Lagrangian mechanics apply in fewer cases than Newtonian mechanics. As we saw, linear drag may look like it has a valid Hamiltonian/Lagrangian, but it doesn't. For quadratic drag or friction due to normal force, one cannot find a suitable trick, and is forced to use Rayleighs dissipation functions which modify the Euler-Lagrange equations. This is not a coincidence: while Newtonian mechanics links kinematics and dynamics by choosing a particular frame, Hamiltonian and Lagrangian mechanics do so by fixing a type of system. It is the implicit knowledge of the type of system that allows us to reconstruct the dynamics just by looking at the kinematics in an unknown frame. What we need to understand, then, is what exactly is this restriction.

1.4 Reversing Hamiltonian mechanics

We now turn our attention to Hamiltonian mechanics and try to understand exactly what types of systems it focuses on. We will find twelve equivalent formulations of Hamiltonian mechanics that link ideas from vector calculus, differential geometry, statistical mechanics, thermodynamics, information theory and plain statistics. The overall result is that Hamiltonian mechanics focuses on systems that are assumed to be deterministic and reversible. We will see how the physical significance of that assumption differs from mathematically naive characterizations.

Mathematical characterizations

To simplify our discussion, we will first concentrate on a single degree of freedom. The first characterization of Hamiltonian mechanics is naturally in terms of the equations

$$\begin{aligned} d_t q &= \partial_p H \\ d_t p &= -\partial_q H. \end{aligned} \tag{HM-1D}$$

We will want to treat phase space as a generic two-dimensional space (i.e. manifold), like we would for a plane in physical space. We will reserve the term coordinate for the position variable q, while we will refer to the collection of position and momentum as state variables and will note them as $\xi^a = [q, p]$. We can now define the displacement field

$$S^a = d_t \xi^a = [d_t q, d_t p] \tag{1.36}$$

which is a vector field that defines the evolution of the system in time. Hamilton's equations, then, can be expressed as

$$\begin{aligned} S^q &= \partial_p H \\ S^p &= -\partial_q H. \end{aligned} \tag{1.37}$$

To bring out the geometric meaning of the equations, we introduce the matrix

$$\omega_{ab} = \begin{bmatrix} \omega_{qq} & \omega_{qp} \\ \omega_{pq} & \omega_{pp} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{SF-1D}$$

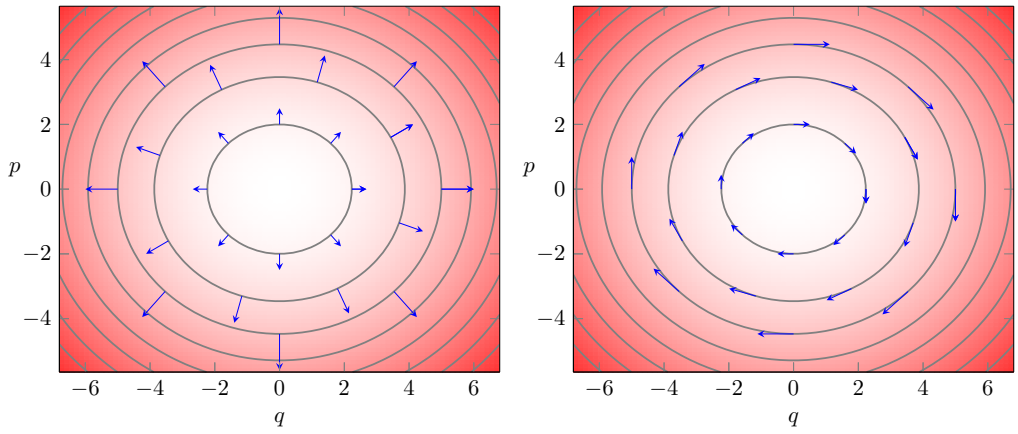


Figure 1.6: The surface plot shows the value of the Hamiltonian for a harmonic oscillator $H = \frac{p^2}{2m} + \frac{1}{2}kq^2$, red means higher value. The lines are the regions at constant energy H. On the left, the gradient of the Hamiltonian is shown. On the right, the displacement field is shown, which is the gradient rotated by a right angle. Note how the displacement is always parallel to the lines at constant energy.

which rotates a vector by a right angle.[13] That is, if $v^a = [v^q, v^p]$, then $v_a = v^b\omega_{ba} = [-v^p, v^q]$.[14] We can rewrite equation HM-1D as

$$S_a = S^b\omega_{ba} = \partial_a H \tag{HM-G}$$

which tells us that the displacement field is the gradient of the Hamiltonian rotated by a right angle. Note that the gradient is perpendicular to the lines at constant energy. Therefore, as we can see in fig. 1.6, a right angle rotation gives us a vector field tangent to those lines, making it geometrically evident that the value of the Hamiltonian is a constant of motion. Condition HM-G is just a re-expression of HM-1D. Though it is already useful, we want to find different mathematical conditions which turn out to be equivalent to the equations.

We start by noting that the displacement field as expressed by 1.37 looks very similar to a curl of H, except that it is a two dimensional version. In vector calculus, a vector field is the curl of another field if and only if its divergence is zero.[15] This holds here as well. First, we can verify that

$$\partial_a S^a = \partial_q S^q + \partial_p S^p = \partial_q\partial_p H - \partial_p\partial_q H = 0. \tag{1.38}$$

Geometrically, this means that the flow of S^a through a closed region is always zero, as shown in fig. 1.7. That is, $\oint (S^q dp - S^p dq) = 0$. Note that, since we are in a two dimensional space, a

[13] The notion of angle is technically ill-defined in phase space, but this slight imprecision makes it easier to get the point across.

[14] The notation is purposely similar to how indexes are raised and lowered in general relativity by the metric tensor $g_{\alpha\beta}$, since ω_{ab} plays a similar geometric role in phase space. One should be careful, however, that ω_{ab} is anti-symmetric (i.e. $\omega_{ab} = -\omega_{ba}$), so it matters which side is contracted. In terms of symplectic geometry, the rotated displacement field S_a corresponds to the interior product of the displacement field with the symplectic form, usually noted as $\iota_S\omega$ or $S \lrcorner \omega$.

[15] We will leave for now topological requirements as they would be a distraction from the overall point.

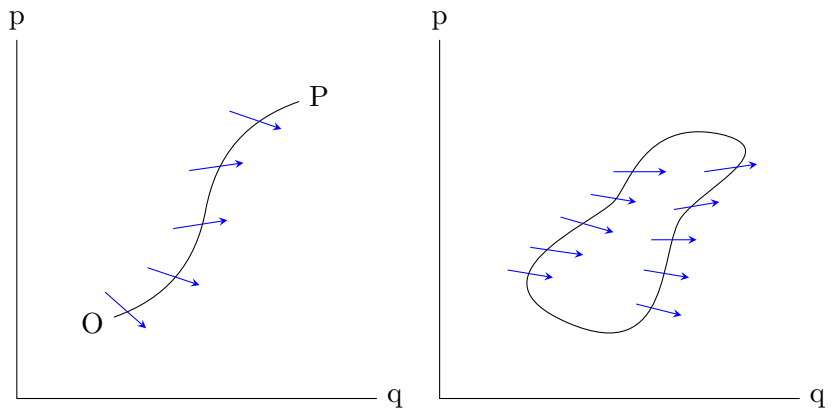


Figure 1.7: The flow of the displacement field S^a through a path, shown on the left, is equal to the difference of the Hamiltonian at the two points $\Delta H = \int_{OP} S^a \times d\xi^b$. The net flow of states through a region (i.e. the flow of the displacement field through the boundary) is zero, as shown on the left. This means that S^a is divergenceless and will admit a stream function, a potential, which corresponds to the Hamiltonian H.

hyper-surface has dimension $n-1 = 2-1 = 1$ and therefore hyper-surfaces are lines. Therefore we have

$$\oint (S^q dp - S^p dq) = \oint (\partial_p H dp + \partial_q H dq) = \oint dH = 0. \tag{1.39}$$

That is, the flow of the displacement field is the line integral of the gradient of H, which is zero over a closed curve.

Conversely, we can see that each divergenceless field in two dimensions admits a stream function H that satisfies HM-1D. Geometrically, we can construct H in the following way. Take a reference state O in phase space and assign $H(O) = 0$. For any other state P, consider the flow of S through any two lines that connect O and P. Given that the flow through the region contoured by those lines must be zero, the flow through each line must be equal. Therefore the flow through a line that connects O and P only depends on the states, it is path independent. We can assign $H(P) = \int_{OP} (S^q dp - S^p dq)$. If we expand the differential of H we have

$$dH = \partial_q H dq + \partial_p H dp = -S^p dq + S^q dp. \tag{1.40}$$

If we equate the components, we recover HM-1D. Geometrically, at least for the one dimensional case, we can understand the difference of the Hamiltonian between two states as the flow of the displacement field between them.

We conclude that the following condition

$$\text{The displacement field is divergenceless: } \partial_a S^a = 0 \tag{DR-DIV}$$

is equivalent to HM-1D. Unlike HM-G, this is a truly different mathematical condition.

Having looked at the flow through a region, we turn our attention to how regions themselves are transported by the evolution. Liouville's theorem states that volumes of phase space

are preserved during Hamiltonian evolution, which in our case will be areas over the $q-p$ plane. To see this, let us review how variables transform, together with infinitesimal volumes:

$$\begin{aligned} \hat{\xi}^a &= \hat{\xi}^a(\xi^b) \\ d\hat{\xi}^a &= \partial_b \hat{\xi}^a d\xi^b \\ d\hat{\xi}^1 \cdots d\hat{\xi}^n &= \left|\partial_b \hat{\xi}^a\right| d\xi^1 \cdots d\xi^n \\ d\hat{q}d\hat{p} &= \begin{vmatrix} \partial_q \hat{q} & \partial_p \hat{q} \\ \partial_q \hat{p} & \partial_p \hat{p} \end{vmatrix} dqdp \end{aligned} \tag{1.41}$$

This tells us that, mathematically, a transformation is volume preserving if the determinant of the Jacobian $\partial_b \hat{\xi}^a$ is unitary. If $\hat{q}$ and $\hat{p}$ represent the evolution of q and p after an infinitesimal time step δt, we have

$$\begin{aligned} \hat{q} &= q + S^q \delta t \\ \hat{p} &= p + S^p \delta t \\ \partial_b \hat{\xi}^a &= \begin{bmatrix} 1 + \partial_q S^q \delta t & \partial_p S^q \delta t \\ \partial_q S^p \delta t & 1 + \partial_p S^p \delta t \end{bmatrix} \\ \left|\partial_b \hat{\xi}^a\right| &= (1 + \partial_q S^q \delta t)(1 + \partial_p S^p \delta t) - \partial_p S^q \, \partial_q S^p \, \delta t^2 = 1 + (\partial_q S^q + \partial_p S^p)\, \delta t + O(\delta t^2). \end{aligned} \tag{1.42}$$

Note that the first order term is proportional to the divergence of the displacement field, therefore the Jacobian determinant is equal to one if and only if the displacement is divergenceless. In other words, condition

$$\text{The Jacobian of time evolution is unitary: } \left|\partial_b \hat{\xi}^a\right| = 1 \tag{DR-JAC}$$

and condition

$$\text{Volumes are conserved through the evolution: } d\hat{\xi}^1 \cdots d\hat{\xi}^n = d\xi^1 \cdots d\xi^n \tag{DR-VOL}$$

are equivalent to DR-DIV. We have found a third and a fourth way to characterize Hamiltonian evolution.

While condition DR-VOL is expressed in terms of areas, similar considerations will work for densities because a density is a quantity divided by an infinitesimal area. In fact densities

$$\left|\partial_b \hat{\xi}^a\right| \hat{\rho}(\hat{\xi}^a) = \rho(\xi^b). \tag{1.43}$$

transform in an equal and opposite way with respect to areas (i.e. the Jacobian determinant is on the other side of the equality). The unitarity of the Jacobian determinant, then, is equivalent to requiring that the density at an initial state is always equal to the density at the corresponding final state. Both areas and densities are transported unchanged by Hamiltonian evolution, as shown in fig. 1.8. Therefore

$$\text{Densities are conserved through the evolution: } \hat{\rho}(\hat{\xi}^a) = \rho(\xi^b) \tag{DR-DEN}$$

is yet another equivalent characterization.

To get a yet different perspective, we can reframe these arguments in terms of ω_{ab} and S_a. Given two vectors v^a and w^a, the area of the parallelogram they form is $v^q w^p - v^p w^q$. This can

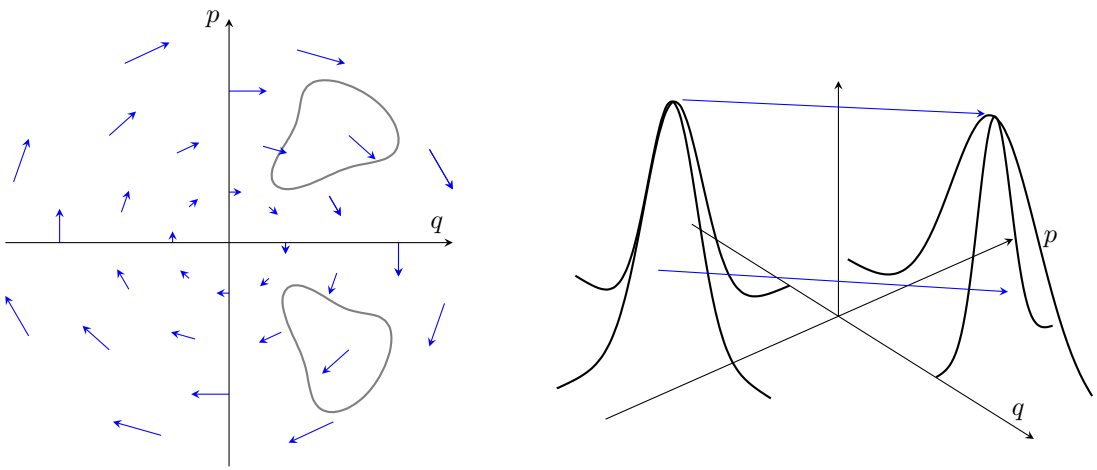


Figure 1.8: On the left side, we see how the displacement field S^a transports areas of phase space to equal areas of phase space. On the right, we see Hamiltonian evolution transports a probability distribution point by point. The value of the probability density remains the same as it moves over phase space.

be rewritten as $v^a \omega_{ab} w^b$, which means we can think of ω_{ab} as a tensor that, given two vectors, returns the area of the parallelogram they form.[16] If we denote $\hat{v}^a = \partial_b \hat{\xi}^a \, v^b$ and $\hat{w}^a = \partial_b \hat{\xi}^a \, w^b$ the transformed vectors, the invariance of the area can be written as

$$v^a \omega_{ab} w^b = \hat{v}^c \omega_{cd} \hat{w}^d. \tag{1.44}$$

Since

$$\hat{v}^c \omega_{cd} \hat{w}^d = v^a \, \partial_a \hat{\xi}^c \omega_{cd} \, \partial_b \hat{\xi}^d \, w^b = v^a \, \hat{\omega}_{ab} w^b \tag{1.45}$$

the previous equivalence means that $\omega_{ab} = \hat{\omega}_{ab}$, that is ω_{ab} remains unchanged. In other words, preserving the area for all possible pairs of vectors is the same as preserving the tensor ω_{ab} that returns the areas. We now see that ω_{ab} plays such an important geometric role that

The evolution leaves ω_{ab} invariant: $\hat{\omega}_{ab} = \omega_{ab}$ (DI-SYMP)

is yet another equivalent characterization of Hamiltonian mechanics.

It is useful to look more closely at the definition of the Poisson bracket

$$\{f, g\} = \partial_q f \, \partial_p g - \partial_p f \, \partial_q g = \begin{vmatrix} \partial_q f & \partial_p f \\ \partial_q g & \partial_p g \end{vmatrix}. \tag{1.46}$$

For a single degree of freedom, the Poisson bracket coincides with the Jacobian determinant, where f and g are the two new variables. It essentially tells us how the volume changes if we change state variables from $[q, p]$ to $[f, g]$. Canonical transformations, then, are those that

[16] More properly, ω_{ab} is a two-form.

do not change the units of area. The Poisson bracket can be expressed[17] as

$$\{f,g\} = -\partial_a f \omega^{ab} \partial_b g = \partial_b g \omega^{ba} \partial_a f \tag{1.47}$$

where

$$\omega^{ab} = \begin{bmatrix} \omega^{qq} & \omega^{qp} \\ \omega^{pq} & \omega^{pp} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \tag{1.48}$$

is the inverse of ω_{ab}. The invariance of the Poisson brackets is equivalent to the invariance of the inverse of ω_{ab}, which is equivalent to DI-SYMP. Therefore

The evolution leaves the Poisson brackets invariant (DI-POI)

is yet another equivalent characterization. So, again, we see how ω_{ab} plays a fundamental geometrical role.

We can also rewrite the flow of the displacement field

$$\int (S^q dp - S^p dq) = \int S^a \omega_{ab} d\xi^b = \int S_b d\xi^b \tag{1.49}$$

as the line integral of the rotated displacement field S_a. We can do that because in two dimensions the flow through a boundary is effectively a line integral along the boundary with the field rotated 90 degrees. This means that the following condition

The rotated displacement field is curl free: $\partial_a S_b - \partial_b S_a = 0$ (DI-CURL)

is equivalent to condition DR-DIV.[18] In fact, we can read equation HM-G as saying that the rotated displacement field is the gradient of the scalar potential H.

We can see that we have found plenty of alternative characterizations of Hamilton's equations HM-1D (or HM-G). Conditions DR-DIV, DR-JAC, DR-VOL and DR-DEN relate more directly to the displacement field S^a, while conditions DI-SYMP, DI-POI and DI-CURL relate more directly to ω_{ab} and the rotated displacement field S_a. Nonetheless, they are all in terms of the mathematical description. While these are useful, the final goal of reverse physics is to find physical assumptions, not just equivalent mathematical definitions. So it is time to step back and try to understand what the math is really about.

Physical characterizations

Let us first reflect on what we just found out: the defining characteristic of Hamiltonian mechanics is not the transport of points, but the transport of areas and densities. If classical Hamiltonian mechanics were really about and only about point particles, there would be no reason for it to be characterized by DR-DIV, DR-VOL or DR-DEN. In fact, there would be

[17] To see how our definitions and notation map to that used in differential geometry, let us define $\partial^a H = \omega^{ab}\partial_b H$. Note that $\partial^a H$ corresponds to the Hamiltonian vector field of H usually noted X_H. The Poisson bracket is usually defined as $\omega(X_f, X_g)$. In our notation this becomes $\partial^a f \omega_{ab} \partial^b g = \omega^{ac}\partial_c f \omega_{ab} \omega^{bd} \partial_d g = \omega^{ac}\partial_c f \delta^d_a \partial_d g = \omega^{ac}\partial_c f \partial_a g$. One can see how the notation mimics the Einstein notation of general relativity and avoids the introduction of ad-hoc symbols.

[18] Those familiar with relativistic electromagnetism will recognize the expression $\partial_a S_b - \partial_b S_a$ as the generalization of the curl. More properly, it is the exterior derivative applied to a one-form.

no reason for the equations of motion HM-1D to be differentiable. Differentiable equations are exactly needed if we need to define the Jacobian, the transport of areas, or of densities defined on those areas. Classical point particles, then, are more aptly conceived not as points, but as infinitesimal regions of phase space, as distributions so peaked that only the mean value is important.

This, in retrospect, matches how classical mechanics is used in practice: planets, cannonballs, pendulums, beads on a wire, all the objects we study with classical mechanics are not point-like objects. They can be considered point-like if their size is negligible compared to the scale of the problem. If the distance between two celestial bodies is smaller than the sum of their radii, the point particle approximation clearly fails. This is also consistent with fluid dynamics and continuum mechanics, where we are literally studying the motion of infinitesimal parts of a material. It is interesting to see echoes of these considerations present in the mathematics.[19]

If we look at physics more broadly, we realize that in statistical mechanics we already have a physical interpretation for volumes of regions in phase space: they represent the number of states. Hamiltonian mechanics, then, maps regions while preserving the number of states. This means that, for each initial state there is one and only one final state, which leads to the following condition:

The evolution is deterministic and reversible. (DR-EV)

Note that by reversible here we mean that given the final state we can reconstruct the initial state. Given that areas measure the number of states, DR-EV is equivalent to DR-VOL, which means this is another characterization of Hamiltonian mechanics. We can also see a connection to DR-DEN. If we assign a density to an initial state, and we claim that all and only the elements that start in that initial state will end in a particular final state, we will expect the density of the corresponding final state to match. That is, if the evolution is deterministic and reversible, it may shuffle around a distribution, but it will never be able to spread it or concentrate it.

This makes us understand, at a conceptual level, why a dissipative system, like a particle under linear drag, is not a Hamiltonian system. A dissipative system will have an attractor: a point or a region to which the system will tend given enough time. This means that, in time, the area around the attractor must shrink, the density will concentrate over the attractor, but this is exactly what Hamiltonian systems cannot do. Therefore Hamiltonian systems cannot have attractors, they cannot be dissipative. By the same argument, they can't have unstable points or regions from which the system always goes away.

What may be confusing is that the motion of a particle under linear drag may seem reversible, in the sense that we are able to, given the final position and momentum, reconstruct the initial values. Mathematically, it maps points one-to-one and would seem to satisfy DR-EV, even though it is not a Hamiltonian system. This is a perfect example of how focusing on just the points leads to the wrong physical intuition. Physically, we would say that a one meter range of position allows for more configurations than a one centimeter range, even though mathematically they have the same number of points. If we understand that states are infinitesimal areas of phase space, we can see that a dissipative system, though it does

[19] We will want to investigate this link in more detail later.

map the center points of infinitesimal areas one-to-one, it does not map the full infinitesimal area one-to-one. In this sense dissipative systems fail to be reversible.

Let that sink in: we found that, if the system is deterministic and reversible, it admits a Hamiltonian, a notion of energy, and that energy is conserved over time. This may seem like a surprising and unexpected result. In retrospect, we can make an argument for it based on familiar physics considerations. If a system is deterministic and reversible it means that its evolution only depends on the state of the system itself. This means that it does not depend on the state of anything else. A system whose evolution does not depend on anything else is an isolated system. Therefore a deterministic and reversible system is isolated, and from thermodynamics we know that an isolated system conserves energy. It should not be surprising, then, that a deterministic and reversible system conserves energy. However, we found that not only does it conserve energy, it defines it. Therefore this link between mechanics and thermodynamics is actually deeper than we may think at first, and we should explore it further.

The idea that a dissipative system is not reversible sounds true on thermodynamic grounds. But thermodynamic reversibility is not the ability to reconstruct the initial state, but rather the existence of a process that can undo the change. Alternatively, a process is thermodynamically reversible if it conserves thermodynamic entropy, which is a more precise characterization.[20] We should not, then, confuse the two notions of reversibility, but we can easily show their relationship. The fundamental postulate of statistical mechanics tells us that the thermodynamic entropy $S = k_B \log W$ is the logarithm of the count of states, which corresponds to volume in phase space. Since the logarithm is a bijective function, conservation of areas of phase space is equivalent to conservation of entropy. Therefore

The evolution is deterministic and thermodynamically reversible (DR-THER)

is yet another characterization of Hamiltonian mechanics.

There is another type of entropy that is also fundamental in both statistical mechanics and information theory: the Gibbs/Shannon entropy $I[\rho(\xi^a)] = -\int \rho \log \rho \, d\xi^1 \cdots d\xi^n$ which is defined for each distribution $\rho(\xi^a)$. Recalling the transformation rules for both volumes 1.41 and densities 1.43, we have

$$\begin{aligned}
I[\rho(\xi^a)] &= -\int \rho(\xi^a) \log \rho(\xi^a) \, d\xi^1 \cdots d\xi^n \\
&= -\int \hat{\rho}(\hat{\xi}^b) \left|\partial_a \hat{\xi}^b\right| \log\left(\hat{\rho}(\hat{\xi}^b) \left|\partial_a \hat{\xi}^b\right|\right) d\xi^1 \cdots d\xi^n \\
&= -\int \hat{\rho}(\hat{\xi}^b) \log\left(\hat{\rho}(\hat{\xi}^b) \left|\partial_a \hat{\xi}^b\right|\right) d\hat{\xi}^1 \cdots d\hat{\xi}^n \\
&= -\int \hat{\rho}(\hat{\xi}^b) \log \hat{\rho}(\hat{\xi}^b) \, d\hat{\xi}^1 \cdots d\hat{\xi}^n - \int \hat{\rho}(\hat{\xi}^b) \log\left|\partial_a \hat{\xi}^b\right| d\hat{\xi}^1 \cdots d\hat{\xi}^n \\
&= I[\hat{\rho}(\hat{\xi}^b)] - \int \hat{\rho}(\hat{\xi}^b) \log\left|\partial_a \hat{\xi}^b\right| d\hat{\xi}^1 \cdots d\hat{\xi}^n .
\end{aligned} \tag{1.50}$$

Information entropy, then, remains constant if and only if the logarithm of the Jacobian determinant is zero, which means the Jacobian determinant is one. Therefore

The evolution conserves information entropy (DR-INFO)

[20] The actual existence of a reverse process is not something that can always be guaranteed.

is equivalent to DR-JAC and is yet another characterization of Hamiltonian mechanics.

The fact that determinism and reversibility is equivalent to conservation of information entropy should not be, in retrospect, surprising. Given a distribution, its information entropy quantifies the average amount of information needed to specify a particular element chosen according to that distribution. If the evolution is deterministic and reversible, giving the initial state is equivalent to giving the final state and therefore the information to describe one or the other must be the same. Determinism and reversibility, then, can be understood as the informational equivalence between past and future descriptions.

Lastly, given that entropy is often associated with uncertainty, it may be useful to understand how Hamiltonian evolution affects uncertainty. Given a multivariate distribution, the uncertainty is characterized by the covariance matrix

$$cov(\xi^a, \xi^b) = \begin{bmatrix} \sigma_q^2 & cov_{q,p} \\ cov_{p,q} & \sigma_p^2 \end{bmatrix}. \tag{1.51}$$

The determinant of the covariance matrix gives us a coordinate independent quantity to characterize the uncertainty. If the distribution is narrow enough, we can use the linearized transformation to see how the uncertainty evolves after an infinitesimal time step δt. We have

$$\left|cov(\hat{\xi}^c, \hat{\xi}^d)\right| = \left|\partial_a \hat{\xi}^c \, cov(\xi^a, \xi^b) \, \partial_b \hat{\xi}^d\right| = \left|\partial_a \hat{\xi}^c\right| \left|cov(\xi^a, \xi^b)\right| \left|\partial_b \hat{\xi}^d\right|, \tag{1.52}$$

which means the uncertainty remains unchanged if and only if the Jacobian is unitary. So

The evolution conserves the uncertainty of peaked distributions (DR-UNC)

is equivalent to DR-JAC and is another characterization of Hamiltonian mechanics.

This connection gives us yet another insight on the nature of determinism and reversibility in physics. Given that all physically meaningful descriptions are finite precision, a system is deterministic and reversible in a physically meaningful sense if and only if the past/future descriptions can be reconstructed/predicted at the same level of precision. This gives us another perspective as to why areas and densities must be conserved.

Assumption of determinism and reversibility

We have found twelve equivalent characterizations that link Hamiltonian mechanics, vector calculus, differential geometry, statistical mechanics, thermodynamics, information theory and plain statistics. Though we only talked about the case of a single degree of freedom, it gives us a much better idea of what systems Hamiltonian mechanics is supposed to describe, those that satisfy the following

Assumption DR (Determinism and Reversibility)**.** *The system undergoes deterministic and reversible evolution. That is, specifying the state of the system at a particular time is equivalent to specifying the state at a future (determinism) or past (reversibility) time.*

We can see how this concept is implemented mathematically: it is not simply a one-to-one map between points. Classical particles should be more properly thought of as infinitesimal regions of phase space. Conceptually, the count of states, the thermodynamic entropy and information entropy are all conserved, and are all equivalent characterizations of determinism

and reversibility. In terms of physical measurement, past and future states are given at the same level of uncertainty. But the most important lesson is that the foundations of classical mechanics are not disconnected from the foundations of all other disciplines we encountered. A full understanding of classical mechanics means understanding those connections as well.

1.5 Multiple degrees of freedom

We have seen how DR is a constitutive assumption for Hamiltonian mechanics, and in fact is equivalent to Hamiltonian mechanics for one degree of freedom. We now turn our attention to the general case, and we will find that DR, by itself, is not enough to recover the equations. We will need an additional assumption, that of the independence of degrees of freedom.

First, let's take Hamilton's equations for multiple degrees of freedom

$$\begin{aligned} d_t q^i &= \partial_{p_i} H \\ d_t p_i &= -\partial_{q^i} H \end{aligned} \tag{HM-ND}$$

and re-express them in terms of generalized state variables. These will be noted as $\xi^a = [q^i, p_i]$ and will span a $2n$-dimensional space (i.e. manifold). The displacement field will be

$$S^a = d_t \xi^a = [d_t q^i, d_t p_i] \tag{1.53}$$

which again is the vector field that defines the evolution of the system in time. Hamilton's equations, then, can be expressed as

$$\begin{aligned} S^{q^i} &= \partial_{p_i} H \\ S^{p_i} &= -\partial_{q^i} H. \end{aligned} \tag{1.54}$$

Similarly to the previous case, let's introduce the following matrix

$$\omega_{ab} = \begin{bmatrix} \omega_{q^i q^j} & \omega_{q^i p_j} \\ \omega_{p_i q^j} & \omega_{p_i p_j} \end{bmatrix} = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \otimes I_n \tag{SF-ND}$$

which performs a 90 degree rotation within each degree of freedom, switching the components between position and momentum. That is, if $v^a = [v^{q^i}, v^{p_i}]$, then $v_a = v^b \omega_{ba} = [-v^{p_i}, v^{q^i}]$.[21] We can rewrite equation HM-ND as

$$S_a = S^b \omega_{ba} = \partial_a H \tag{1.55}$$

which notationally is the same as HM-G. The insight that the displacement field is equal to the gradient of H rotated 90 degrees still applies, except there are now multiple ways, in principle, to do that rotation. It is only the one defined by ω_{ab} that works.

[21] For those versed in symplectic geometry, $v^a \omega_{ab}$ are the components of the one-form $\omega(v, \cdot)$. However, we are not going to call it a one-form as that assumes that the whole object is a map from a vector field to a scalar field, and we do not know whether that is the correct physical understanding. In other words, we want simply to understand what the quantities are doing without being tied, as much as possible, to a particular way to frame it. Full reverse engineering of differential geometry will be done in a later chapter, once the physics we need to describe is clear.

Conditions DR-DIV, DR-JAC, DR-VOL and DR-DEN are still satisfied and equivalent to each other. In fact, the divergence of the displacement field is zero

$$\partial_a S^a = \partial_{q^i} S^{q^i} + \partial_{p_i} S^{p_i} = \partial_{q^i}\partial_{p_i} H - \partial_{p_i}\partial_{q^i} H = 0 \tag{1.56}$$

and the Jacobian is unitary

$$\begin{aligned}
\hat{q}^i &= q^i + S^{q^i}\,\delta t \\
\hat{p}_i &= p_i + S^{p_i}\,\delta t \\
\partial_b \hat{\xi}^a &= \begin{bmatrix} \delta^i_j + \partial_{q^j} S^{q^i}\,\delta t & \partial_{p_j} S^{q^i}\,\delta t \\ \partial_{q^j} S^{p_i}\,\delta t & \delta^j_i + \partial_{p_j} S^{p_i}\,\delta t \end{bmatrix} \\
\left|\partial_b \hat{\xi}^a\right| &= \left|\delta^i_j + \partial_{q^j} S^{q^i}\,\delta t\right| \left|\delta^j_i + \partial_{p_j} S^{p_i}\,\delta t\right| - \left|\partial_{p_j} S^{q^i}\,\delta t\right| \left|\partial_{q^j} S^{p_i}\,\delta t\right| \\
&= 1 + \left(\partial_{q^i} S^{q^i} + \partial_{p_i} S^{p_i}\right)\delta t + O(\delta t^2)
\end{aligned} \tag{1.57}$$

since the first-order term is again the divergence. The Jacobian is still the multiplicative factor between past/future areas (and densities), and therefore they are conserved even in the case of multiple degrees of freedom.

However, these conditions are not equivalent to HM-ND. The displacement field S^a has $2n$ components and is therefore specified by $2n$ functions. Conditions DR-DIV, DR-JAC, DR-VOL and DR-DEN specify the same single constraint, bringing down to $2n-1$ the number of independent components. The choice of Hamiltonian provides another constraint, leaving $2n-2$ choices undetermined. In the single degree of freedom case, $n=1$, no choices are left, and therefore the displacement field is fully constrained. In the general case, however, this is not enough to fully characterize the evolution. Therefore HM-ND implies DR-DIV, DR-JAC, DR-VOL and DR-DEN, but the converse is not true.

Let's see what happens to condition DI-SYMP, the invariance of ω in the general case. We have

$$\begin{aligned}
\hat{\omega}_{ab} &= \partial_a \hat{\xi}^c \omega_{cd} \partial_b \hat{\xi}^d \\
&= \left(\delta^c_a + \partial_a S^c \delta t\right) \omega_{cd} \left(\delta^d_b + \partial_b S^d \delta t\right) \\
&= \omega_{ab} + \left(\partial_a S^c \omega_{cb} + \omega_{ad}\partial_b S^d\right)\delta t + O(\delta t^2) \\
&= \omega_{ab} + \left(\partial_a (S^c \omega_{cb}) + \partial_b (S^d \omega_{ad})\right)\delta t + O(\delta t^2) \\
&= \omega_{ab} + \left(\partial_a (S^c \omega_{cb}) - \partial_b (S^d \omega_{da})\right)\delta t + O(\delta t^2).
\end{aligned} \tag{1.58}$$

Therefore, the invariance of ω_{ab} is equivalent to

$$\partial_a (S^c \omega_{cb}) - \partial_b (S^c \omega_{ca}) = 0. \tag{1.59}$$

In terms of the rotated displacement field S_a we have the more compact form

$$\partial_a S_b - \partial_b S_a = 0. \tag{1.60}$$

This tells us that the rotated displacement field S_a is curl free, which is the same condition as DI-CURL, therefore DI-CURL and DI-SYMP are equivalent conditions also in the general case.

Note that Hamilton's equations state that the rotated displacement field is the gradient of the Hamiltonian, and therefore

$$\partial_a S_b - \partial_b S_a = \partial_a(S^c\omega_{cb}) - \partial_b(S^c\omega_{ca}) = \partial_a\partial_b H - \partial_b\partial_a H = 0, \tag{1.61}$$

which simply verifies that the curl of the gradient is zero. Conversely, if S_a is curl-free, then it admits a scalar potential H such that

$$S_a = S^b\omega_{ba} = \partial_a H \tag{1.62}$$

which recovers Hamilton's equations. Therefore HM-ND, DI-CURL and DI-SYMP are equivalent.

The relationship between Poisson brackets and ω^{ab} is the same in the general case, therefore DI-POI and DI-SYMP are equivalent as well.

To sum up, in the general case HM-ND, HM-G, DI-SYMP, DI-POI and DI-CURL are all equivalent and therefore full characterizations of Hamiltonian mechanics in the general case. These imply DR-DIV, DR-JAC, DR-VOL and DR-DEN, which are all equivalent to one another, but weaker conditions that cannot recover Hamiltonian mechanics in full. For the second set of conditions, we already have an intuitive geometrical picture: the net flow of the displacement within a region of phase space is zero, volumes are preserved and so are densities. We need to build a stronger geometrical intuition for the first set, which is actually the more fundamental one.

Condition DI-SYMP tells us that $v^a\omega_{ab}w^b$ is a conserved quantity, no matter what vectors v^a and w^b we choose. In the case of a single degree of freedom, this represented the area of the parallelogram formed by the two vectors, which was also the volume of the region. In the general case, we still have two vectors, but the situation is a bit more complicated.

We can gain an understanding by looking at the outer product decomposition for ω_{ab} we saw in SF-ND. This tells us that what happens within a degree of freedom is different from what happens across degrees of freedom. If we pick a single degree of freedom $1 \leq x \leq n$ and two vectors $v = v^q e_{q^x} + v^p e_{p_x}$ and $w = w^q e_{q^x} + w^p e_{p_x}$ that stretch along that degree of freedom, then we have

$$v^a\omega_{ab}w^b = v^q w^p - v^p w^q. \tag{1.63}$$

That is, within each degree of freedom, ω_{ab} computes the area of the parallelogram. Since ω_{ab} is conserved, parallelograms within any degree of freedom will be mapped to parallelograms of the same size.

If we pick two different DOFs x and y and two corresponding vectors $v = v^q e_{q^x} + v^p e_{p_x}$ and $w = w^q e_{q^y} + w^p e_{p_y}$, then we have

$$v^a\omega_{ab}w^b = 0. \tag{1.64}$$

This defines a notion of orthogonality between different degrees of freedom. Since ω_{ab} is conserved, this notion of orthogonality is preserved during the evolution: orthogonal degrees of freedom are mapped to orthogonal degrees of freedom.

Those familiar with general relativity and/or Riemannian geometry may gain more insight by the following analogy. In those cases, the metric tensor g_{ij} defines the geometry by defining the scalar product between vectors. That is, given two vectors v^i and w^j,

$v^i g_{ij} w^j = |v||w| \cos\theta_{vw}$. Therefore the metric tensor defines the length and angles for vectors. In Cartesian coordinates, the metric tensor is a unitary matrix of the same dimension of the space. The form ω_{ab} does something in some sense similar and in some sense different. It defines areas within degrees of freedom and angles between them. Hamiltonian evolution preserves these areas and angles.

If areas and orthogonality are preserved, then volumes are preserved as well. The volume of a parallelepiped formed by parallelograms on orthogonal degrees of freedom will simply be the product of the areas of the parallelograms. Therefore we can understand why Hamiltonian mechanics satisfies DR-VOL. We can also understand why DR-VOL is not enough to recover Hamiltonian mechanics. An evolution could stretch one degree of freedom while shrinking another by the same amount. The total volume would remain the same, even though the area in each degree of freedom wouldn't. For example, take the system of equations:

$$\begin{aligned} d_t q^1 &= S^{q^1} = \frac{p_1}{m} \\ d_t p_1 &= S^{p_1} = -b p_1 \\ d_t q^2 &= S^{q^2} = \frac{p_2}{m} \\ d_t p_2 &= S^{p_2} = b p_2 \end{aligned} \tag{1.65}$$

The first degree of freedom is a particle under linear drag, while the second is a particle accelerated (not decelerated) proportionally to its momentum by the same coefficient. We can verify that

$$\partial_a S^a = \partial_{q^1} \frac{p_1}{m} + \partial_{p_1}(-b p_1) + \partial_{q^2} \frac{p_2}{m} + \partial_{p_2}(b p_2) = -b + b = 0 \tag{1.66}$$

the divergence is zero and therefore DR-DIV and DR-VOL are satisfied. However

$$\partial_{q^1} S_{p_1} - \partial_{p_1} S_{q^1} = \partial_{q^1} S^{q^1} \omega_{q^1 p_1} - \partial_{p_1} S^{p_1} \omega_{p_1 q^1} = \partial_{q^1} \frac{p_1}{m}(1) - \partial_{p_1}(-b p_1)(-1) = -b. \tag{1.67}$$

The curl of S_a, then, is not zero, DI-CURL is not satisfied, nor are DI-SYMP and HM-ND. The system is not Hamiltonian precisely because we are not preserving the areas within each independent DOF: the first is shrunk and the second stretched.

Now that we have a more precise understanding of the mathematics and the geometry, we should turn to the physics. Note that all the previous physical conditions DR-EV, DR-THER, DR-INFO and DR-UNC are equivalent to DR-VOL and DR-JAC. Therefore determinism and reversibility is clearly a constitutive assumption of Hamiltonian mechanics in the general case, but it cannot be the only one. Ideally, we would like to find a condition that is independent of DR. However, we saw that DI-SYMP implies DR-VOL, therefore the mathematics does not already give us two independent conditions we can map to the physics.

This is an important aspect to understand for reverse physics: the mapping between physical and mathematical conditions need not necessarily be one to one. A single mathematical condition can map to multiple physical ones, or the same physical condition can map to multiple mathematical ones. We saw before that determinism and reversibility forces the evolution map to be both bijective and volume preserving. Mathematically, these are two independent conditions. We can have a bijection that is not volume preserving (e.g. a linear transformation that stretches one side) or a volume preserving map that is not bijective (e.g. a map from

$\mathbb{R}$ to $\mathbb{R}$ that maps all rationals to 0 while leaving all the irrationals the same). Yet, a physically meaningful deterministic and reversible map must do both. Here we have the opposite: Hamiltonian mechanics implies determinism and reversibility, but is also implying at least another physical condition, and we need to understand which and whether it is physically independent.

Let's start from what we have already established: the phase space volume quantifies the number of states in the region. It stands to reason that the area on each degree of freedom identifies the number of configurations for that degree of freedom. Therefore, given two vectors, $v = v^q e_{q^x} + v^p e_{p_x}$ and $w = w^q e_{q^x} + w^p e_{p_x}$, constrained on a single degree of freedom x, the area of the parallelogram they identify, $v^q w^p - v^p w^q$, quantifies the number of configurations. Therefore ω_{ab} returns the number of configurations within each degree of freedom. What about between degrees of freedom?

As we saw, the conserved volume represents the total number of states, and it is the product of those degrees of freedom that are orthogonal. In terms of ω_{ab}, x and y are orthogonal if $\omega_{q^x q^y} = \omega_{q^x p_y} = \omega_{p_x q^y} = \omega_{p_x p_y} = 0$. In this case, the volume is simply the product of the phase-space areas on the x and y degrees of freedom. This means that the total number of states is the product of the configurations of each degree of freedom. Physically, it means that a configuration choice for x does not constrain the configurations for y. This means that the degrees of freedom are independent.

Given that there are different notions of independence, let us go through an example. Suppose we have a rabbit farm and describe its state with the number of males and females. These two variables are independent: if we say there are 231 females it doesn't, in principle, tell us anything about the number of males. Now, we may expect the population of both sexes to be about equal, and we may even find that in most rabbit farms that is the case, but this does not describe something about the nature of the variables themselves: it describes the nature of rabbit farms. Chicken farms, for example, would be predominantly females, as those are the ones that lay eggs.

Now we could choose to describe the rabbit farm with another set of variables: the number of females and the total number of rabbits. In this case, the variables are not independent. If we find that there are 231 females, it tells us, in principle, that there must be at least 231 rabbits. Conversely, if we find that there are 231 rabbits, there can only be up to 231 females. This dependence is not a feature of the rabbit farms. It does not just happen to be that there are no farms where the number of female rabbits exceeds the total number of rabbits. There can't be one.

This type of independence is very different from the notion of independence in terms of statistics and probability. The latter is in terms of whether the probability distribution factorizes. That is, if $P(f, m)$ is the probability that a particular rabbit farm has f females and m males, the distributions of males $P(m)$ and females $P(f)$ are independent if $P(f, m) = P(f)P(m)$.

We therefore have two notions of independence. One is on the variables themselves and whether they can allow (i.e. whether we can measure) different combinations. One is on the probability distribution we may have in a particular case and whether it factorizes. The orthogonal directions in phase space, then, are independent in the first, stronger, sense. The degrees of freedom themselves are independent, regardless of what probability distribution one may put on top.

One may ask whether there is a link between the two, and in fact there is. Going back to

our rabbits, we can easily see that, given any distribution $P(f)$ on the females, we can choose any distribution $P(m)$ on the males and set $P(f,m) = P(f)P(m)$. However, this does not happen for the total number. Suppose we chose $P(f)$ and we wanted to find a $P(f+m)$ such that $P(f, f+m) = P(f)P(f+m)$. The probability of the total number of rabbits could not change based on the number of females. If the probability of having 231 females is non-zero, then the probability of having less than 231 total rabbits in that case must be zero. But since we want the probability of having less than 231 rabbits independent of the number of females, then it must be zero for all cases. That is, the probability $P(f+m)$ must be zero for all numbers smaller than the greatest value of f such that $P(f) \neq 0$. If there is no such greatest value of f, for example f follows a geometric distribution, no $P(f+m)$ can exist that is independent of $P(f)$.

The conclusion is that only independent variables can support independent distributions.[22] It should be clear that this observation is something that goes beyond the physical underpinning of Hamiltonian mechanics: it is something that applies to any variable to which we want to assign a probability distribution. As such, we do not want to expand the scope too much at this point, though clearly we will need to explore this more in full.

Here we limit ourselves to concluding that the following four conditions

The system is decomposable into independent DOFs	(IND-DOF)
The system allows statistically independent distributions over each DOF	(IND-STAT)
The system allows informationally independent distributions over each DOF	(IND-INFO)
The system allows peaked distributions where the uncertainty is the product of the uncertainty on each DOF	(IND-UNC)

are equivalent. The first means that the count of states factorizes, the second that probability distributions can factorize, the third that the information entropy can sum, and the fourth that the determinant of the covariance matrix can factorize. Since only independent variables can support statistically independent distributions, IND-DOF is equivalent to IND-STAT. Statistical independence of random variables coincides with independence of information entropy, therefore IND-STAT is equivalent to IND-INFO. The uncertainty for peaked distributions factorizes if and only if the joint distribution is the product of independent distributions, therefore IND-STAT is equivalent to IND-UNC.

Clearly these conditions are independent from DR. We can imagine a deterministic and reversible system that cannot be broken into separate independent degrees of freedom, and we can imagine a system that can be broken into separate independent degrees of freedom that does not evolve deterministically or reversibly. The question is whether assuming independent degrees of freedom and deterministic and reversible evolution is enough to recover Hamiltonian mechanics.

The first thing to check, then, is whether we have enough constraints to recover ω_{ab}. Assuming that the system can be broken up into independent degrees of freedom, we must be

[22] Formally, let $(\Omega, \mathcal{F}, P)$ be a probability space, let $X : \Omega \to E_X$ and $Y : \Omega \to E_Y$ be two random variables and $Z : \Omega \to E_X \times E_Y$ be their joint random variable (i.e. $Z(\omega) = (X(\omega), Y(\omega))$. Then X and Y are independent in this stronger sense if $Z(\Omega) = X(\Omega) \times Y(\Omega)$ and are statistically independent if the cumulative distribution function $F_Z(x,y) = F_X(x)F_Y(y)$ factorizes. Alternatively, they are independent if the σ-algebra generated by the joint distribution Z is the product of the σ-algebras generated by X and Y, and are statistically independent if the σ-algebras generated by X and Y are independent in the standard probability sense.

able to define the count of configurations for each degree of freedom, and independent degrees of freedom must be orthogonal. The fact that ω_{ab} will return zero if it acts on directions belonging to independent degrees of freedom is really telling us that ω_{ab} counts not just configurations, but independent configurations. This, in retrospect, makes sense. But, so far this seems to restrict ω_{ab} to be

$$\omega_{ab} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \otimes \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix}. \tag{1.68}$$

The fact that ω_{ab} must return the area in each degree of freedom constrains the left matrix of the outer product to the one above. The fact that ω_{ab} needs to return zero across independent degrees of freedom constrains the right matrix of the outer product to be diagonal. However, there is nothing, at this point, that seems to constrain the area of each DOF to map to the same count of configurations. Naturally, we could simply rescale conjugate momentum in each DOF to homogenize the count, but this would be an arbitrary freedom. Is there something forcing all the a_i to be the same?

Note that q^i and p_i are not the only variables that form independent degrees of freedom. If we take two independent DOFs x and y, $x+y$ and $x-y$ will also form independent degrees of freedom. That is, ω_{ab} defines orthogonality for all DOFs, not just those of a particular basis. Changing x and y to $x+y$ and $x-y$ will effectively apply a rotation on the diagonal matrix, which will remain diagonal only if the coefficients on the diagonal are the same.

This tells us that if we want ω_{ab} to properly capture the independence of linear combinations of independent DOFs, the diagonal matrix must have the same coefficient. This coefficient represents the freedom we have in choosing the units of omega with respect to the units of everything else. In SI units, by convention, the product between q^i and p_i is in $J \cdot s$ (i.e. the same units of h and angular momentum) and we set the coefficient to 1. Therefore expressing the number of configurations with the same units for all DOFs is not an extra constraint, but it is necessary to keep track of the dependency relationship for all degrees of freedom.

Therefore the other constitutive assumption of Hamiltonian mechanics is

Assumption IND (Independent DOFs)**.** *The system is decomposable into independent degrees of freedom. That is, the variables that describe the state can be divided into groups that have independent definition, units and count of states.*

This assumption leads to conditions IND-DOF, IND-STAT, IND-INFO and IND-UNC, which we saw implies the existence of a form ω_{ab} that defines the independence of DOFs together with the count of independent configurations for each DOF. Conversely, assuming HM-ND means defining an ω_{ab} such that IND is satisfied. The last question is whether DR and IND are enough to recover HM-ND.

As we saw, IND by itself means the existence of the counting form ω_{ab}. Meanwhile, DR by itself means the conservation of the total number of states, the volume. These two mathematical conditions, by themselves, do not lead to Hamiltonian mechanics. We need that the counting form itself is conserved, meaning that DOF independence and the configuration count is preserved. This boils down to the following question: does it make sense, on physical

grounds, to have a deterministic and reversible evolution that takes a system decomposable into independent degrees of freedom and turns it into a system that is no longer decomposable? More specifically, can deterministic and reversible evolution take two independent degrees of freedom and break their independence?

We should remind ourselves that independence here is not statistical independence. Clearly, Hamiltonian evolution can add correlations, and therefore evolve the product of two independent distributions into something that is no longer factorizable. This is not the question at hand. The notion of independence on the table is the one that tells us that all the combinations of configurations are at least possible. Recall the example of the rabbits: number of females and total rabbits are not independent because we cannot have more females than rabbits. This is the type of independence we are interested in. Can deterministic and reversible evolution lose this type of independence?

The answer is, as you may expect, no. This is easily seen in the finite case. Suppose we have two integer variables, $1 \leq x \leq 3$ and $1 \leq y \leq 3$. If they are independent, we have a total of 9 distinct cases. If the evolution is deterministic and reversible, we will still have 9 distinct cases, which means the variables must remain independent. Note that we may introduce a correlation between x and y. But still, we need 9 total cases.

The issue here is that the case of independent variables is maximal as it posits that all combinations of configurations are possible. Therefore, the only way to make variables no longer independent is to decrease the number of distinct cases, which cannot happen during deterministic and reversible evolution. The same must happen in the infinite or the continuous case, because finite ranges must still be comparable with finite sizes which must hold the same property. Therefore when we take both physical assumptions IND and DR, the physics tells us that independence of degrees of freedom must be preserved, which means preserving ω_{ab} as well. In other words:

Insight 1.69. *Hamiltonian mechanics is exactly the deterministic and reversible evolution of a system decomposable into a finite collection of independent degrees of freedom.*

This conclusion is yet another example of why just looking at the math is not enough. Two physical conditions taken separately may each impose one mathematical condition, but it is not necessarily true that imposing them together will only impose the conjunction of the two mathematical conditions. More than a problem with math in general, we believe it is an indication that the math we currently use is not the "physically correct" one as it does not seem to be capturing the entirety of the physical conditions.

Note that, in principle, we could ask the evolution to preserve the independence of DOFs without requiring DR. As we saw, fixing all independent DOFs fixes ω_{ab} up to a scalar factor, which could change during the evolution. The volume would stretch or shrink depending on the factor, stretching or shrinking each DOF by the same amount. An example of this would be a particle under linear drag in three dimensions. Since a faster moving object will be subjected to a greater frictional force, a spread in momentum Δp will become smaller in time, and will tend to zero as time increases. Given that the friction coefficient is the same for all directions, all degrees of freedom will shrink at the same rate. If we understand the volume as entropy, this tells us that the only way we can add or remove entropy to/from a system while preserving the independence of the degrees of freedom is by dividing that entropy contribution equally among each DOF. In other words, preservation of DOF independence

gives us a sort of equipartition of entropy change. It is again striking to find these connections between disciplines at such a basic level.

To conclude, we now have a mathematically and physically precise way to characterize Hamiltonian evolution. We had found that Hamiltonian mechanics did not apply to all cases, and now we know exactly to which cases it applies: systems described by finitely many independent degrees of freedom undergoing deterministic and reversible evolution.

1.6 Reversing differential topology

In the previous sections, we saw that elements of differential topology and differential geometry started appearing: we employed a generalization of the curl and we used the form ω_{ab} to lower indexes, much like one does in general relativity with the metric tensor $g_{\alpha\beta}$. Since we will use these tools more and more, we should take a detour and understand the physical significance of the tools themselves. We will end up with a generalized notion of integral and of differential operations that work with an arbitrary number of dimensions. We will also conclude that to reach the full potential of reverse physics, we need to apply the same techniques not just to the equations themselves, but to the mathematical tools we use to formulate them.

The main reason we are forced to abandon vector calculus in favor of differential topology and geometry is that vector calculus works in three dimensions but does not generalize to an arbitrary dimensional space. Special and general relativity, for example, live on a four-dimensional space-time; Hamiltonian and Lagrangian mechanics live in phase space, which can have an arbitrarily large number of degrees of freedom. Similarly to what we have done with the equations, we will start from the expressions of vector calculus, see their limitations, and construct generalized ones. We warn the reader who is already familiar with differential topology that this process will give us notation and concepts that are slightly different from what is used by mathematicians. We will discuss these differences at the end of the section.

The first tool we need to generalize is that of line, surface and volume integrals. For example, the mass m within a region V can be understood as the sum of the contributions of a mass density ρ within each infinitesimal volume dV:

$$m(V) = \iiint_V \rho dV. \tag{1.70}$$

Similarly, the magnetic flux Φ through a surface Σ can be understood as the sum of the contributions of the magnetic field $\vec{B}$ over each infinitesimal $d\vec{\Sigma}$:

$$\Phi(\Sigma) = \iint_\Sigma \vec{B} \cdot d\vec{\Sigma}. \tag{1.71}$$

Lastly, the work W over a path γ can be understood as the sum of the contributions of a force field $\vec{f}$ over each infinitesimal segment $d\vec{\gamma}$:

$$W(\gamma) = \int_\gamma \vec{f} \cdot d\vec{\gamma}. \tag{1.72}$$

Note the pattern: the functionals W, Φ and m all take a region of space while $\vec{f}$, $\vec{B}$ and ρ act on infinitesimal regions. However the pattern is not completely consistent. In the line integral case, we have the product between a vector representing the force and a vector representing the displacement along the line. In the surface integral case we have the product between a

pseudo-vector representing the magnetic force and a pseudo-vector representing the normal to the surface element. In the volume integral case we have the product between a pseudo-scalar representing the density and a pseudo-scalar representing the volume of the infinitesimal region. Each operation is slightly ad-hoc. Moreover, a surface has a single perpendicular direction only in three dimensions. In four dimensions, for example, there are multiple different perpendiculars to the same plane. Lastly, they all require a notion of product between vectors, an inner product, which in differential geometry is only defined on Riemannian spaces, those that define a metric tensor. That is, those spaces in which angles and distances are well defined. In physics we are so used to working with spaces that have an inner product that it may seem that all spaces provide one, but that is not the case. In phase space, there is no way to compare differences in position with differences in momentum, therefore we do not have a metric to take a scalar product; there is no overall notion of distance and angle. In general, if we imagine the space that represents all possible outcomes of a blood test, this will form a manifold as each test can be fully described by a finite set of continuous quantities. In this space, there is no natural notion of distance and angles between directions, there is no notion of geometry. As we will see, the notion of integral, the idea of a quantity that can be understood as the sum of infinitely many infinitesimally small contributions, does not require either a particular number of dimensions or a notion of distance and angle.

Suppose we understood $\vec{f}$, $\vec{B}$ and ρ not as above, but as maps that for each infinitesimal region return an infinitesimal contribution dW, $d\Phi$ and dm. We could simply write

$$\begin{aligned} W(\gamma) &= \int_\gamma dW = \int_\gamma f(d\gamma) \\ \Phi(\Sigma) &= \iint_\Sigma d\Phi = \iint_\Sigma B(d\Sigma) \\ m(V) &= \iiint_V dm = \iiint_V \rho(dV). \end{aligned}$$

This pattern is straightforward and more easily generalized. We call these functions of infinitesimal regions k-forms, where k is the dimensionality of the infinitesimal region they take as an argument. The force, in this notation, is a one-form (or covector) as it takes one dimensional infinitesimal regions (i.e. vectors); the magnetic field is a two-form; the density is a three-form. We can also say that a scalar field, like the temperature, is a zero-form, as it takes points, zero dimensional objects.

Since k-forms act over infinitesimal regions, they will have some key properties. First, note that each infinitesimal region can be understood as a parallelepiped, and a parallelepiped is fully identified by its sides. Therefore, a k-form can be understood as acting on a set of infinitesimal displacements, the sides of the parallelepiped, whose number matches the dimensionality of the form. A one-form will take one displacement, a two-form two displacements and so on. Second, as they are linear functions of the infinitesimal regions, they will also be linear functions of the vectors that define these infinitesimal regions. Lastly, all forms must be anti-symmetric because switching the order of the sides does not change the parallelepiped, but it changes its orientation.

We can write displacements and forms in terms of components and basis elements

$$\begin{aligned} dP &= dx^i e_i \\ f &= f_i e^i \\ B &= B_{ij} e^i \otimes e^j \\ \rho &= \rho_{ijk} e^i \otimes e^j \otimes e^k. \end{aligned} \tag{1.73}$$

The anti-symmetry means that switching the indexes introduces a minus sign. For example, $B_{ij} = -B_{ji}$ and $\rho_{ijk} = -\rho_{jik}$. Therefore, our integrals can be written as

$$\begin{aligned} W(\gamma) &= \int_\gamma f_i \, dx^i \\ \Phi(\Sigma) &= \iint_\Sigma B_{ij} \, dx_1^i \, dx_2^j \\ m(V) &= \iiint_V \rho_{ijk} \, dx_1^i \, dx_2^j \, dx_3^k. \end{aligned} \tag{1.74}$$

Each k-form, then, is a fully anti-symmetric covariant tensor, with one index for each dimension of the form. In this view, the B^x component of the magnetic field, then, becomes the B_{yz} component. That is, instead of being the component that gives us the magnetic flux along the x direction, it is the component that gives us the flux through the yz plane. Similarly, ρ_{xyz} is better understood not as the value of the mass density at a point, but rather the value of the density over the xyz volume. The indexes make unit dependency apparent as well: B_{yz} must be in units of magnetic flux divided by the units of y and z; ρ_{xyz} must be in units of mass divided by the units of x, y and z; in spherical coordinates, $B_{r\theta}$ must be in units of magnetic flux divided by units of r and θ. In fact, one can argue that these tools are here exactly to keep track of the units of the components.

We can go further, and rewrite the integrals in terms of parametrizations of the surfaces and the components of the forms along said parametrization. We have

$$\begin{aligned} W(\gamma) &= \int_\gamma f_i dx^i = \int_{a_u}^{b_u} f_i \partial_u x^i du = \int_{a_u}^{b_u} f_u du \\ \Phi(\Sigma) &= \iint_\Sigma B_{ij} \, dx_1^i \, dx_2^j = \int_{a_u}^{b_u} \int_{a_v}^{b_v} B_{ij} \, \partial_u x_1^i du \, \partial_v x_2^j dv = \int_{a_u}^{b_u} \int_{a_v}^{b_v} B_{uv} du dv \\ m(V) &= \iiint_V \rho_{ijk} \, dx_1^i \, dx_2^j \, dx_3^k = \int_{a_u}^{b_u} \int_{a_v}^{b_v} \int_{a_w}^{b_w} \rho_{ijk} \, \partial_u x_1^i du \, \partial_v x_2^j dv \, \partial_w x_3^k dw \\ &= \int_{a_u}^{b_u} \int_{a_v}^{b_v} \int_{a_w}^{b_w} \rho_{uvw} du dv dw. \end{aligned} \tag{1.75}$$

These expressions are useful to write the integrals in terms of an integral of k variables. This is particularly useful if the surfaces possess different symmetries than the forms, which allows one to use the parametrization appropriately.

We now have expressions that are easier to generalize. If we denote S^k the space of all the k-dimensional subregions of an n-dimensional manifold, a k-functional $F_k : S^k \to \mathbb{R}$ is an additive functional that takes a k-surface σ^k and returns a number. This can be expressed as

$$\begin{aligned} F_k(\sigma^k) &= \int_{\sigma^k} \omega_k(d\sigma^k) = \int_{\sigma^k} \omega_{i_1 i_2 \cdots i_k} \, dx_1^{i_1} \, dx_2^{i_2} \cdots dx_k^{i_k} \\ &= \int_{a_{u_1}}^{b_{u_1}} \int_{a_{u_2}}^{b_{u_2}} \cdots \int_{a_{u_k}}^{b_{u_k}} \omega_{u_1 u_2 \cdots u_k} \, du_1 \, du_2 \cdots du_k, \end{aligned} \tag{1.76}$$

which is the integral of a k-form $\omega_k : V^k \to \mathbb{R}$. Here V is the space of vectors, of infinitesimal displacements, and an infinitesimal k-surface $d\sigma^k$ is defined by an ordered set of k vectors, therefore is an element from the Cartesian product $V^k = V \times V \times \cdots \times V$. The k-form takes an infinitesimal k-surface $d\sigma^k$ and returns a number. We now have a way to express local objects and integrals in a generalized way.

In the spirit of reverse physics, we ask: what are the physical assumptions that are required to use these mathematical objects? Note that we do not measure mass density ρ directly: we measure mass m within a finite region V and the size of the finite region V, and we calculate the mass density by dividing the first by the second. The mass density ρ is the limit for which the region V shrinks to a point.[23] The same happens with the other quantities. We do not measure a force f directly, but rather the work that a force performs, for example, by deforming a spring. In other words, we measure the finite value of the functional F over a finite region σ, and the form ω returns the limit of F for infinitesimal regions. Note that it is crucial for F to be additive, that is, if σ_1 and σ_2 are two disjoint regions, $F(\sigma_1 \cup \sigma_2) = F(\sigma_1) + F(\sigma_2)$. Physically, it means that the contribution of one sub-region is independent from the other, because if this isn't the case, we cannot assign a unique value to each region. While this may seem a relatively harmless assumption, it may not hold in general. For example, suppose we have a system distributed in space. Given the mass-energy equivalence, the mass is an additive functional only if the interaction between the parts can be neglected. In fact, if the interaction energy at the boundary is so large that it is of the same scale of the mass within the regions, then it is not true that the total mass is simply the mass within each region. Therefore, if we write the mass density ρ, we have implicitly assumed that the interaction energy between parts can be neglected. While this is going to be true in a large number of cases, it is something we have to keep in mind. The functional F, in fact, is a more general physical object as it may exist even though infinitesimal additivity fails.

If the functional F respects the limits, in the sense that small variations of the region result in small variations of the value of the functional, then we can express the functional F as the integral of a form ω.[24] Note that the standard mathematical definitions are inverted with respect to what makes physical sense. Mathematically, we first define the form and then its integration, and we may worry whether the integral exists or diverges; physically, we first define the functional of finite regions and worry about the existence of an infinitesimal limit, the form, under additional physical assumptions. This is important because, as we said before, it tells us that it is the functional that survives when the assumption fails, not the form. It also means that, if we want to get a robust physical intuition, we should concentrate on the functionals, the finite objects, rather than the forms, the infinitesimal objects.

We therefore reach the following:

Insight 1.77. *Differential k-forms represent the infinitesimal contributions of an infinitesimally additive quantity F_k that depends on a k-dimensional surface.*

The issue, however, is that while we have a sense that, mutatis mutandi, an infinitesimally additive functional corresponds to a differential form, we would need to show that differential

[23] This is essentially the definition of the Radon-Nikodym derivative used in measure theory.

[24] Mathematically, one needs to be precise as to what these small variations are, and what the space of regions is. Effectively, we need to define what it means for k-surfaces to be differentiable. This is more in the scope of physical mathematics than reverse physics.

forms cannot represent anything else. We currently do not have such an argument, though we suspect it should exist. One may first note that no measurement can really be conducted at a point, and therefore it has to extend, at least in principle, over a k-dimensional region. Therefore we could argue that all measurements are functionals, additive or not. This means that we would need a map between the value of the components of the form at all points and the value of the functional over all regions. To reach our goal, we would need to show that this map can be established only if the functional is infinitesimally additive, which may not be a strong enough condition by itself. While we are already implicitly assuming the map to be differentiable, we could impose further requirements on the measurement functional. We could impose locality, in the sense that changes of the form within a region must change the value of the functional for said region; it would also mean that changes outside of the region do not affect measurements in the region. While we do not know whether these requirements are sufficient, these are the types of questions we would need to answer in order to have a fully physically meaningful treatment of differential forms.

Having generalized the idea of integration, let us now turn to differential operators such as gradient, curl and divergence. Suppose that T is a scalar field, like the temperature. Then we have the following relationship

$$\int_\gamma \vec{\nabla} T \cdot d\vec{\gamma} = T(B) - T(A). \tag{1.78}$$

That is, the integral of the gradient of T along γ gives us the difference of T evaluated at the endpoints, the boundary of the line. If $\vec{f}$ is a vector field, like the force, we have

$$\iint_\Sigma \vec{\nabla} \times \vec{f} \cdot d\vec{\Sigma} = \int_{\gamma=\partial\Sigma} \vec{f} \cdot d\vec{\gamma} = W(\gamma) \tag{1.79}$$

That is, the integral of the curl of $\vec{f}$ over a surface Σ is equal to the line integral of $\vec{f}$ over the boundary. If $\vec{B}$ is a pseudo-vector field, like the magnetic field, we have

$$\iiint_V \vec{\nabla} \cdot \vec{B}\, dV = \iint_{\Sigma=\partial V} \vec{B} \cdot d\vec{\Sigma} = \Phi(\Sigma). \tag{1.80}$$

That is, the integral of the divergence of $\vec{B}$ over a region V is equal to the surface integral of $\vec{B}$ over the boundary. Note the pattern: the integral of the differential operator applied to the bulk becomes the integral of the original object over the boundary.

Let us understand how this works in terms of functionals. Given the temperature $T(P)$, we can construct a line functional that for each line returns the difference in temperature at the endpoints. Given the work line functional $W(\gamma)$, we can construct a surface functional that for each surface returns the work needed to go around the contour. Given the magnetic flux surface functional $\Phi(\Sigma)$, we can construct a volume functional that for each volume returns the magnetic flux over the boundary. In general, suppose we have a k-functional $F_k : S^k \to \mathbb{R}$. To each k-surface σ^k we can associate a quantity $F_k(\sigma^k)$. Now, suppose we are given a $(k+1)$-surface σ^{k+1}. While F_k cannot act on σ^{k+1}, the boundary $\partial\sigma^{k+1}$ is a k-surface, therefore we can evaluate $F_k(\partial\sigma^{k+1})$. Therefore, we can define the $(k+1)$-functional ∂F_k such that $\partial F_k(\sigma^{k+1}) \mapsto F_k(\partial\sigma^{k+1})$, which we call the exterior functional of F_k.[25] Since the

[25] Mathematically, we would have to prove that ∂F_k is a functional. Again, these mathematical details are left for the physical mathematics section. For now, we are more interested in the conceptual understanding.

exterior functional is an additive functional, it will have a corresponding form that acts on the infinitesimal region. Conceptually, the gradient of T is the one-form that corresponds to the boundary functional of T; the curl of $\vec{f}$ is the two-form that corresponds to the boundary functional of the work W, the line integral of $\vec{f}$; the divergence of $\vec{B}$ is the three-form that corresponds to the boundary functional of the magnetic flux Φ, the surface integral of $\vec{B}$. The problem is, again, that the gradient, curl and divergence are not expressed in a way that is easy to generalize.

Given that all three operators are in terms of ∇, which in components is written ∂_i, we would like to write something along the lines of

$$\begin{aligned}\partial F_k(\sigma^{k+1}) &= \int_{\sigma^{k+1}} \partial_{i_0} \wedge \omega_{i_1 i_2 \cdots i_k} dx^{i_0} dx^{i_1} dx^{i_2} \cdots dx^{i_k} \\ &= \int_{\partial\sigma^{k+1}} \omega_{i_1 i_2 \cdots i_k} dx^{i_1} dx^{i_2} \cdots dx^{i_k} = F_k(\partial\sigma^{k+1}).\end{aligned} \tag{1.81}$$

The operation $\wedge$, which we call exterior product, must be such that we recover something consistent to the previous operations. That is, we need

$$\begin{aligned}\partial_i \wedge T &= \partial_i T \\ \partial_i \wedge F_j &= \partial_i F_j - \partial_j F_i \\ \partial_i \wedge B_{jk} &= \partial_i B_{jk} + \partial_j B_{ki} + \partial_k B_{ij}\end{aligned} \tag{1.82}$$

These would be the expressions we need to recover the gradient, curl and divergence respectively. Let us study them to see the pattern. Fist of all, each expression takes a k-form and returns a $(k+1)$-form by adding a derivation along each index. Given that we have a derivation for each index, the number of terms matches the number of indexes of the final form, which is $k+1$. Each term changes index by taking a cyclic permutation. Recall that the forms are anti-symmetric, therefore each permutation of two indexes introduces a minus sign. A cyclic permutation of $k+1$ elements corresponds to k pair swaps. If $k+1$ is odd, then, each cyclic permutation will correspond to an even number of sign switches, which cancel out. The pattern, then, generalizes in the following way

$$\begin{aligned}\partial_{i_0} \wedge \omega_{i_1 i_2 \cdots i_k} &= \partial_{i_0}\omega_{i_1 i_2 \cdots i_k} + (-1)^k \partial_{i_1}\omega_{i_2 \cdots i_k i_0} + (-1)^{2\cdot k}\partial_{i_2}\omega_{i_3 \cdots i_k i_0 i_1} + \cdots \\ &\quad + (-1)^{k\cdot k}\partial_{i_k}\omega_{i_0 i_1 \cdots i_{k-1}} \\ &= \sum_{j=0}^{k} (-1)^{j\cdot k} \partial_{i_{j \bmod k+1}} \omega_{i_{j+1 \bmod k+1} i_{j+2 \bmod k+1} \cdots i_{j+k \bmod k+1}}.\end{aligned} \tag{1.83}$$

The use of $\bmod k+1$ in the generalized expression makes sure that the index jumps from i_k to i_0. This gives us a fully anti-symmetric tensor which matches the gradient, curl and divergence in the simple cases.[26] This operation is called the exterior derivative.

While we have the expression for the exterior derivative, we would like to understand why and how the expression works. Geometrically, we can imagine integrating along a parallelepiped, which becomes

$$\int_{\sigma^{k+1}} (\partial \wedge \omega_k)(d\sigma^{k+1}) = \int_{\sigma^{k+1}} \partial_{i_0} \wedge \omega_{i_1 i_2 \cdots i_k}\, dx_0^{i_0}\, dx_1^{i_1} \cdots dx_k^{i_k}$$

[26] In principle, we could write the expression in different equivalent ways. Here we used cyclic permutations as it gives a more elegant expression.

$$
\begin{aligned}
&= \int_{a_{u_0}}^{b_{u_0}} \int_{a_{u_1}}^{b_{u_1}} \cdots \int_{a_{u_k}}^{b_{u_k}} (\partial_{u_0}\omega_{u_1u_2\cdots u_k} + (-1)^k \partial_{u_1}\omega_{u_2\cdots u_k u_0} \\
&+ \cdots + (-1)^{k\cdot k}\partial_{u_k}\omega_{u_0u_1\cdots u_{k-1}})du_0\, du_1\, du_2 \cdots du_k \\
&= \int_{a_{u_1}}^{b_{u_1}} \cdots \int_{a_{u_k}}^{b_{u_k}} \int_{a_{u_0}}^{b_{u_0}} du_0 \partial_{u_0}\omega_{u_1u_2\cdots u_k}\, du_1\, du_2 \cdots du_k \\
&+ (-1)^k \int_{a_{u_0}}^{b_{u_0}} \int_{a_{u_2}}^{b_{u_2}} \cdots \int_{a_{u_k}}^{b_{u_k}} \int_{a_{u_1}}^{b_{u_1}} du_1 \partial_{u_1}\omega_{u_2\cdots u_k u_0}\, du_0\, du_2 \cdots du_k \\
&+ \cdots + (-1)^{k\cdot k} \int_{a_{u_0}}^{b_{u_0}} \int_{a_{u_1}}^{b_{u_1}} \cdots \int_{a_{u_k}}^{b_{u_k}} du_k \partial_{u_k}\omega_{u_0u_1\cdots u_{k-1}}\, du_0\, du_1 \cdots du_{k-1} \\
&= \left[\int_{a_{u_1}}^{b_{u_1}} \cdots \int_{a_{u_k}}^{b_{u_k}} \omega_{u_1u_2\cdots u_k}\, du_1\, du_2 \cdots du_k \right]_{a_{u_0}}^{b_{u_0}} \\
&+ (-1)^k \left[\int_{a_{u_0}}^{b_{u_0}} \int_{a_{u_2}}^{b_{u_2}} \cdots \int_{a_{u_k}}^{b_{u_k}} \omega_{u_2\cdots u_k u_0}\, du_0\, du_2 \cdots du_k \right]_{a_{u_1}}^{b_{u_1}} \\
&+ \cdots + (-1)^{k\cdot k} \left[\int_{a_{u_0}}^{b_{u_0}} \int_{a_{u_1}}^{b_{u_1}} \cdots \int_{a_{u_{k-1}}}^{b_{u_{k-1}}} \omega_{u_0u_1\cdots u_{k-1}}\, du_0\, du_1 \cdots du_{k-1} \right]_{a_{u_k}}^{b_{u_k}} \\
&= \int_{\partial\sigma^{k+1}} \omega_k(d\partial\sigma^{k+1}).
\end{aligned}
$$

What happens is that each direction of integration will match a derivative in the same direction, and therefore will reduce to the integration of ω on opposing sides of the parallelepiped. This happens for each direction, and therefore the whole integral will reduce to the integration of ω on the surface of the parallelepiped. We have verified that equation 1.81, which is known as the generalized Stokes theorem, indeed works and it includes, as particular cases, the gradient theorem 1.78, the curl theorem 1.79 and the divergence theorem 1.80.

Another aspect of vector calculus is given by the following identities

$$
\begin{aligned}
&\vec{\nabla} \times \vec{\nabla} T = 0 \\
&\vec{\nabla} \cdot \vec{\nabla} \times \vec{f} = 0.
\end{aligned}
\tag{1.84}
$$

To generalize them, we note that the exterior product $\wedge$ is an anti-commutative and associative operation. Therefore we have

$$
\partial_i \wedge \partial_j \wedge \omega_{l_1 l_2 \cdots l_k} = (\partial_i \partial_j - \partial_j \partial_i) \wedge \omega_{l_1 l_2 \cdots l_k} = 0 \wedge \omega_{l_1 l_2 \cdots l_k} = 0. \tag{1.85}
$$

In other words, the exterior derivative applied twice returns zero, no matter on what form it is applied. Given that the curl is the exterior derivative applied to one forms and the gradient is the exterior derivative applied to zero forms, the fact that the curl of the gradient is zero is simply an application of the more general property. The same applies for the divergence of the curl. As we saw, mathematically the property is easy enough to verify, but we get no insight into its meaning. To understand the geometrical significance of the generalized relationship, recall that the exterior derivative of the form is associated with the exterior functional. We should, then, look at what happens when we construct the exterior functional of an exterior functional. We have

$$
\partial\partial F_k(\sigma^{k+2}) = \partial F_k(\partial\sigma^{k+2}) = F_k(\partial\partial\sigma^{k+2}) = F_k(\varnothing) = 0. \tag{1.86}
$$

In words, the exterior of the exterior functional of F_k equals F_k applied to the boundary of the boundary. However, the boundary of a boundary is always the empty set, and any functional applied to the empty set must be zero. In fact, since functionals are additive, we must have

$$F_k(\sigma^k) = F_k(\sigma^k \cup \varnothing) = F_k(\sigma^k) + F_k(\varnothing). \tag{1.87}$$

Therefore, we can see that the identity $\partial \wedge \partial \wedge \omega = 0$ for every form ω is ultimately a direct consequence that $\partial\partial U = \varnothing$ for every set U.[27] This shows how studying differential relationships in terms of finite functionals can give a more geometrically meaningful picture.

The last element of vector calculus we need to generalize is the idea of potentials. That is,

$$\begin{aligned} \vec{\nabla} \times \vec{f} = 0 &\Rightarrow \vec{f} = \vec{\nabla} V \\ \vec{\nabla} \cdot \vec{B} = 0 &\Rightarrow \vec{B} = \vec{\nabla} \times \vec{A}. \end{aligned} \tag{1.88}$$

These are generalized by the following formula

$$\partial_i \wedge \omega_{l_1 l_2 \cdots l_k} = 0 \Rightarrow \omega_{l_1 l_2 \cdots l_k} = \partial_{l_1} \wedge \theta_{l_2 \cdots l_k}. \tag{1.89}$$

That is, if the exterior derivative of a k-form is zero, then there exists a $(k-1)$-form whose exterior derivative is the original k-form.[28]

To sum up, we have generalized the idea of integration over k-dimensional submanifolds of an n-dimensional space, which leads to the idea of k-functionals over finite regions and k-forms over the infinitesimal ones; we have seen that the finite functionals are physically more fundamental than the infinitesimal forms. We have seen how k-functionals induce $(k+1)$-functionals by acting on the boundaries of the $k+1$ dimensional regions. We have seen that the exterior derivative gives us the form associated to the exterior functionals, and that this operation generalizes the notion of gradient, curl and divergence. While this does not exhaust all that can be done in differential topology and differential geometry, this is enough for what we need to use in the following sections.

Those already familiar with differential topology will have noticed that our notation and definitions do not quite match the ones typically used in math textbooks. The issue is that said notation and definitions do not match what we need to physically capture, and therefore it would be a mistake to employ them. Let us briefly see why. First of all, in the context of differential topology vectors are defined as directional derivatives. That is, a vector $v = v^i \partial_i$ is an operator that acts on scalar functions. A velocity, which in physics we think of as a vector, is not a directional derivative. In differential topology, a covector $\theta = \theta_i dx^i$ is defined

[27] Note that we are talking about the boundary of a manifold, not the boundary of a set in the topological sense.

[28] Technically, closed forms (i.e. those whose exterior derivative is zero) are not necessarily exact forms (i.e. those that are the exterior derivative of another form). This is true only on contractible regions (i.e. those regions that can be continuously shrunk to a point, that do not have holes). While this is a subtle mathematical point, we can understand it by looking at the corresponding functionals. An exact form corresponds to a functional that returns zero for any closed surface. A closed form, however, is guaranteed to return zero only on closed surfaces that are contractible, that can be continuously shrunk to a point. For example, the functional associated with a closed form may return non-zero over a closed surface that encloses a hole, something the functional associated with an exact form cannot do. Therefore, all exact forms are closed, but not all closed forms are exact. However, if we restrict ourselves to a contractible region, the two definitions are the same.

as a map from a vector to a scalar number. Conjugate momentum, whose components change as a covector, is not a map. In differential topology, a differential dx^i is a covector, and it is the exterior derivative of the coordinate x^i. This means that differentials are maps such that $dx^i(\partial_{x^j}) = \delta^i_j$. In physics, this is not how we think of differentials and integration. In fact, consider the expression $\int f_i dx^i$. To write it in terms of invariant objects, we would have $f = f_i e^i$ and $dx = dx^i e_i$, where e^i and e_i are the co-basis and basis respectively, and therefore $e^i(e_j) = \delta^i_j$. So we obtain $\int f(dx) = \int f_i e^i(dx^j e_j) = \int f_i dx^i = \int df$. Therefore, in physics, the differential dx is more properly thought of as a vector, where the dx^i are the contravariant components, while f as a map from a vector dx to the differential df, which is what is integrated. Therefore the differential is the vector, while the force is the covector. This is in contrast to the use in differential topology. Moreover, in differential topology there is a notion of a single tangent space where all vectors live, which is not compatible with the idea of units. Consider the basis ∂_i. Given that coordinates are expressed in different units, we cannot simply sum derivatives along different directions. For example, in polar coordinates ∂_r may have units of inverse meters while ∂_θ of inverse radians. Worst of all, a directional derivative is taken with respect to a parameter, which will also have units and physical dimension. For example, if v represents a velocity, the components v^i would also depend on the units of space and time. This would mean that units of vectors, the tangent space of a manifold, depend not only on the physical dimensions of the space, but also on the physical dimensions of all possible parameters along which we may want to define a directional derivative. This would mean that the tangent space is not definable only in terms of the units of the manifold itself, and therefore is not defined just in terms of the manifold itself.

The takeaway message here is the following: the mathematical tools we inherit from mathematics are not necessarily designed to capture the physical relationships we need to capture. Mathematicians only care about formal definitions, regardless of what, or if, they represent physically. In physics we do not have this luxury. If we want to have meaningful physical theories, which is ultimately the goal of reverse physics, we need to revisit the mathematical tools we use to formulate them.

1.7 Reversing Lagrangian mechanics

Now that we have a good geometric and physical feel for Hamiltonian mechanics, and that we have a general understanding of what the tools of differential topology describe physically, we will analyze Lagrangian mechanics more in detail. Conceptually, we already know that Lagrangian mechanics is Hamiltonian mechanics plus assumption KE. We will see that the flow of states in phase space admits a vector potential and the Lagrangian is the scalar product between that potential and the displacement along the path. The principle of least action is geometrically equivalent to asking for paths that are always tangent to the displacement field. Moreover, the principle of least action is better understood as a property of Hamiltonian evolution, and assumption KE is only required to express the product between potential and displacement in terms of velocity instead of momentum.

Kinematic assumption revisited

As we saw in section 1.2, Lagrangian mechanics is the subset of Hamiltonian mechanics for which assumption KE is valid, which means Lagrangian mechanics is equivalent to assuming

DR, IND and KE. For the first two assumptions, we found a host of equivalent mathematical, geometric and physical formulations. Let's see what can we find for KE.

First of all, let us summarize the conditions we have already found. Hamilton's equations always impose that the velocity is a function of the state variables:

$$v^i = d_t q^i = \partial_{p_i} H. \tag{1.90}$$

At fixed position, the Jacobian of the transformation is therefore the Hessian of the Hamiltonian

$$\partial_{p_i} v^j = \partial_{p_i} \partial_{p_j} H. \tag{1.91}$$

Therefore condition

> At every position, the relationship between momentum and velocity is invertible and differentiable (WKE-INV)

and condition

> At every point, the Hessian of the Hamiltonian is non-singular (hyperregularity of H): $|\partial_{p_i} \partial_{p_j} H| \neq 0$ (WKE-HYP)

are equivalent to each other. The non-singularity of the Hessian can also be understood as strict monotonicity of the derivative along momentum. Therefore

> The Hamiltonian is twice differentiable and concave (or convex) in momentum (WKE-CONC)

is another equivalent condition.

Note that the relationship between velocity and momentum is not just invertible, but differentiable as well. This may seem like an additional condition from KE, but recall that the dynamics is not in terms of points, but rather cells of phase space and density distributions. In order to express those geometric elements in terms of the kinematic variables, we must make sure that the Jacobian determinant of the transformation from state variables to kinematic variables is well-defined and non-zero. We have

$$|J| = \begin{vmatrix} \partial_{q^j} x^i & \partial_{p_j} x^i \\ \partial_{q^j} v^i & \partial_{p_j} v^i \end{vmatrix} = \begin{vmatrix} \delta^i_j & 0 \\ \partial_{q^j} v^i & \partial_{p_j} v^i \end{vmatrix} = \left|\delta^i_j\right| \left|\partial_{p_j} v^i\right| - |0| \left|\partial_{q^j} v^i\right| = \left|\partial_{p_j} v^i\right|. \tag{1.92}$$

The Jacobian determinant of the change of variables, then, coincides with the Jacobian determinant of the relationship between velocity and momentum at constant position. Therefore the following conditions are all equivalent to each other and to condition WKE-INV.

> The Jacobian of the transformation between state variables and kinematic variables is non-singular. (WKE-NSIN)
>
> Densities over phase space can be expressed in terms of position and velocity: $\rho(x^i, v^j)|J| = \rho(q^i, p_j)$. (WKE-DEN)
>
> Areas and volumes in phase space can be expressed in kinematic variables: $dx^1 \cdots dx^n \, dv^1 \cdots dv^n = |J| \, dq^1 \cdots dq^n \, dp_1 \cdots dp_n$. (WKE-VOL)
>
> The symplectic form ω_{ab} can be expressed in kinematic variables. (WKE-SYMP)
>
> The displacement field S^a can be expressed in kinematic variables. (WKE-DISP)

The insight is that differentiability between state variables and kinematic variables is required to be able to express the objects we used to characterize the deterministic and reversible dynamics. The only physical requirement, then, is to be able to express the dynamics in terms of the kinematics, which is exactly what assumption KE already requires.

On the physical meaning of the Lagrangian

A key problem is to understand what the Lagrangian represents physically. It is often introduced as the difference between kinetic and potential energy. However, this does not work in general. Take the Lagrangian for a particle with charge $\mathfrak{q}$ under an electromagnetic field with electric potential V and magnetic potential A_i.

$$L = \frac{1}{2}m|v^i|^2 + \mathfrak{q}v^i A_i - \mathfrak{q}V. \tag{1.93}$$

The first term is clearly kinetic energy, the last clearly potential energy, but what about the middle term? It depends on velocity, so it would appear to be a kinetic term, but it also depends on the potential. It's both and neither. The characterization of Lagrangian as difference between kinetic and potential energy, then, works for some systems but not in general and it is best abandoned.

Another problem in understanding what the Lagrangian represents is that it is not unique. Now, a similar problem exists for the Hamiltonian, in the sense that we can sum an arbitrary constant to any Hamiltonian without changing the equations of motion. However, the degeneracy for a Lagrangian is far worse. For example, let $f(x^i, t)$ be an arbitrary function of position and time. We can set

$$L' = L + \partial_{x^i} f(x^i, t)v^i + \partial_t f(x^i, t). \tag{1.94}$$

We can see how the Euler-Lagrange equations 1.3 are affected

$$\begin{aligned} 0 = \partial_{x^i} L' - d_t \partial_{v^i} L' &= \partial_{x^i} L + \partial_{x^i}\partial_{x^j} f v^j + \partial_{x^i}\partial_t f - d_t\left(\partial_{v^i} L + \partial_{x^i} f\right) \\ &= \partial_{x^i} L - d_t\partial_{v^i} L + \partial_{x^i}\partial_{x^j} f v^j + \partial_{x^i}\partial_t f - \partial_{x^j}\partial_{x^i} f d_t x^j - \partial_t\partial_{x^i} f d_t t \\ &= \partial_{x^i} L - d_t\partial_{v^i} L + \partial_{x^i}\partial_{x^j} f v^j + \partial_{x^i}\partial_t f - \partial_{x^j}\partial_{x^i} f v^j - \partial_t\partial_{x^i} f \\ &= \partial_{x^i} L - d_t\partial_{v^i} L. \end{aligned} \tag{1.95}$$

That is, the equations of motion given by L' are the same as the ones given by L. Therefore the actual value, or the difference of values, of the Lagrangian is physically meaningless. This makes the question even more puzzling: what is the Lagrangian?

The extended phase space

Given that we have a good understanding of Hamiltonian mechanics, let's work on the equations that link the two. We have

$$\begin{aligned} L &= p_i v^i - H \\ &= p_i d_t q^i - H d_t t \\ &= \begin{bmatrix} p_i & 0 & -H \end{bmatrix} \begin{bmatrix} d_t q^i \\ d_t p_i \\ d_t t \end{bmatrix}. \end{aligned} \tag{1.96}$$

The Lagrangian, then, can be understood as the scalar product of two vectors. The second one is the displacement along the path, however it is the displacement not just in position and momentum, but over time as well. The correct setting to understand Lagrangian mechanics,

then, is phase space extended by the time variable. If we redefine $\xi^a = [q^i \; p_i \; t]$ to include the time variable, we have

$$S^a = d_t \xi^a = \begin{bmatrix} d_t q^i & d_t p_i & d_t t \end{bmatrix}. \tag{1.97}$$

We can then define the new vector θ_a in terms of its components

$$\theta_a = \begin{bmatrix} p_i & 0 & -H \end{bmatrix}. \tag{1.98}$$

We also need to generalize ω_{ab} to the extended phase space. Looking at equation HM-G, the idea is to put the gradient of the Hamiltonian in the time component. If we set

$$\omega_{ab} = \begin{bmatrix} \omega_{q^i q^j} & \omega_{q^i p_j} & \omega_{q^i t} \\ \omega_{p_i q^j} & \omega_{p_i p_j} & \omega_{p_i t} \\ \omega_{t q^j} & \omega_{t p_j} & \omega_{tt} \end{bmatrix} = \begin{bmatrix} 0 & \delta^i_j & \partial_{q^i} H \\ -\delta^j_i & 0 & \partial_{p_i} H \\ -\partial_{q^j} H & -\partial_{p_j} H & 0 \end{bmatrix}, \tag{SF-EPS}$$

we have

$$\begin{aligned} S^a \omega_{a q^j} &= S^{q^i} \omega_{q^i q^j} + S^{p_i} \omega_{p_i q^j} + S^t \omega_{t q^j} \\ &= -S^{p_j} - S^t \partial_{q^j} H = -S^{p_j} - \partial_{q^j} H = 0 \\ S^a \omega_{a p_j} &= S^{q^i} \omega_{q^i p_j} + S^{p_i} \omega_{p_i p_j} + S^t \omega_{t p_j} \\ &= S^{q^j} - S^t \partial_{p_j} H = S^{q^j} - \partial_{p_j} H = 0 \\ S^a \omega_{at} &= S^{q^i} \omega_{q^i t} + S^{p_i} \omega_{p_i t} + S^t \omega_{tt} \\ &= S^{q^i} \partial_{q^i} H + S^{p_i} \partial_{p_i} H \\ &= \partial_{p_i} H \partial_{q^i} H - \partial_{q^i} H \partial_{p_i} H = 0. \end{aligned} \tag{1.99}$$

This means that, on the phase space extended by time, Hamilton's equations become

$$S_a = S^b \omega_{ba} = 0. \tag{HM-EPS}$$

Note that while the position and momentum components of S_a still perform a rotation, the time component does not. So we can't understand S_a as a rotated displacement. Recall that $v^a \omega_{ab} w^b = v_b w^b$ quantified the number of states in the parallelepiped formed by v^a and w^b. If $v_b w^b = 0$, then the parallelepiped does not identify states on an independent DOF. For each vector v^a, the covector v_a identifies the direction that forms an independent DOF with v^a. If we only have position and momentum, we can see that you get the direction rotated by ninety degrees along each DOF. If we extend phase space with time, time does not add a new independent degree of freedom. In fact, the direction given by the displacement field S^a should give us no new independent states: there should be no direction v^b in phase space such that $S^a \omega_{ab} v^b \neq 0$. In other words, $S_b = S^a \omega_{ab}$ must be zero and this is what equation HM-EPS says.

Potential of the flow - 1 DOF case

We wrote the Lagrangian $L = \theta_a d_t \xi^a$ in terms of θ_a and $d_t \xi^a$, but we have yet to understand what θ_a is. If we compare it to ω_{ab}, we note that the first has the Hamiltonian as a component,

while the second has its derivative. Given that ω_{ab} is anti-symmetric, it is a two-form, we may want to calculate the anti-symmetrized derivative of θ_a, the exterior derivative. We have

$$\begin{aligned}
\partial_a\theta_b - \partial_b\theta_a &= \begin{bmatrix} \partial_{q^i}\theta_{q^j} - \partial_{q^j}\theta_{q^i} & \partial_{q^i}\theta_{p_j} - \partial_{p_j}\theta_{q^i} & \partial_{q^i}\theta_t - \partial_t\theta_{q^i} \\ \partial_{p_i}\theta_{q^j} - \partial_{q^j}\theta_{p_i} & \partial_{p_i}\theta_{p_j} - \partial_{p_j}\theta_{p_i} & \partial_{p_i}\theta_t - \partial_t\theta_{p_i} \\ \partial_t\theta_{q^j} - \partial_{q^j}\theta_t & \partial_t\theta_{p_j} - \partial_{p_j}\theta_t & \partial_t\theta_t - \partial_t\theta_t \end{bmatrix} \\
&= \begin{bmatrix} \partial_{q^i}p_j - \partial_{q^j}p_i & \partial_{q^i}0 - \partial_{p_j}p_i & \partial_{q^i}(-H) - \partial_t p_i \\ \partial_{p_i}p_j - \partial_{q^j}0 & \partial_{p_i}0 - \partial_{p_j}0 & \partial_{p_i}(-H) - \partial_t 0 \\ \partial_t p_j - \partial_{q^j}(-H) & \partial_t 0 - \partial_{p_j}(-H) & \partial_t(-H) - \partial_t(-H) \end{bmatrix} \\
&= \begin{bmatrix} 0-0 & 0-\delta^j_i & -\partial_{q^i}H - 0 \\ \delta^i_j - 0 & 0-0 & -\partial_{p_i}H - 0 \\ 0 + \partial_{q^j}H & 0 + \partial_{p_j}H & -\partial_t H + \partial_t H \end{bmatrix} \\
&= \begin{bmatrix} 0 & -\delta^j_i & -\partial_{q^i}H \\ \delta^i_j & 0 & -\partial_{p_i}H \\ \partial_{q^j}H & \partial_{p_j}H & 0 \end{bmatrix}.
\end{aligned} \tag{1.100}$$

This means that the form ω_{ab} is minus the exterior derivative of θ_a:

$$\begin{aligned}
\omega_{ab} &= -(\partial_a\theta_b - \partial_b\theta_a) = -\partial_a \wedge \theta_b \\
\theta_a &= \begin{bmatrix} p_i & 0 & -H \end{bmatrix}.
\end{aligned} \tag{1.101}$$

In other words, ω_{ab} has a null exterior derivative and θ_a is its potential.

Given that the extended phase space for a single degree of freedom is three dimensional, we can use standard vector calculus to gain more understanding. In this case, we have

$$\begin{aligned}
\theta_a &= \begin{bmatrix} p & 0 & -H \end{bmatrix} \\
\omega_{ab} &= \begin{bmatrix} 0 & 1 & \partial_q H \\ -1 & 0 & \partial_p H \\ -\partial_q H & -\partial_p H & 0 \end{bmatrix} = \begin{bmatrix} 0 & S^t & -S^p \\ -S^t & 0 & S^q \\ S^p & -S^q & 0 \end{bmatrix} = \epsilon_{abc}S^c
\end{aligned} \tag{1.102}$$

where ϵ_{abc} is the fully anti-symmetric Levi-Civita symbol which returns the sign of the permutation of the variables (i.e. $\epsilon_{qpt} = \epsilon_{ptq} = \epsilon_{tqp} = 1$ while $\epsilon_{tpq} = \epsilon_{pqt} = \epsilon_{qtp} = -1$). We then have the following relationship

$$\begin{aligned}
\epsilon_{abc}S^c = \omega_{ab} &= -\partial_a \wedge \theta_b = -(\partial_a\theta_b - \partial_b\theta_a) \\
\vec{S} &= -\vec{\nabla} \times \vec{\theta}_b
\end{aligned} \tag{1.103}$$

Apart from the minus, this is the same relationship we would have between the magnetic field B^i and its vector potential A_i. Therefore the same concepts from vector calculus apply: S^a is divergenceless, admits a vector potential, and the flow of the displacement over a closed surface is zero. We can understand this as assumption DR implemented in the extended phase space. Moreover, we have the following relationship:

$$\int_\Sigma \omega_{ab} dx^a_1 dx^b_2 = \int_\Sigma \epsilon_{abc}S^c dx^a_1 dx^b_2 \tag{1.104}$$

That is, the integral of the form ω_{ab} corresponds to the surface integral of the displacement field S^a.[29] In phase space extended by time, then, ω_{ab} quantifies both the number of states over a surface and the flow of S^a through the surface. Under assumption DR this works because we have one and only one evolution for each state, therefore quantifying the number of evolutions that intersect a given surface is the same as quantifying the number of states that flow through them.

One striking feature of the vector potential θ_a is that it is fully characterized by a single arbitrary component, the time component $-H$. The magnetic vector potential A_i, instead, is characterized by two arbitrary components. What are the exact physical conditions that allow that to happen?

Suppose that S^a is a divergenceless field, then we can write it as minus the curl of a potential

$$\theta_a = \begin{bmatrix} \theta_q & \theta_p & \theta_t \end{bmatrix}. \tag{1.105}$$

Vector potentials are defined up to the gradient of an arbitrary function (i.e. up to a gauge), since $\nabla \times (\theta + \nabla f) = \nabla \times \theta$. We can choose f such that $\partial_p f = -\theta_p$, and therefore we can set, without loss of generality, the momentum component to zero

$$\theta_a = \begin{bmatrix} \theta_q & 0 & \theta_t \end{bmatrix}. \tag{1.106}$$

So far, this procedure can be applied to any potential of a divergenceless field. However, the displacement field S^a is particular in that its time component $S^t = d_t t = 1$ is unitary. That is, states flow at a uniform rate in time. Therefore we must have

$$S^t = -(\partial_q \theta_p - \partial_p \theta_q) = \partial_p \theta_q = 1. \tag{1.107}$$

Integrating the relationship we find that $\theta_q = p + c(q,t)$, where $c(q,t)$ is an arbitrary function. We can choose $c(q,t) = 0$ without loss of generality since that arbitrariness corresponds to a choice of gauge. In fact, we can choose $f(q,t)$ such that $\partial_q f = -c(q,t)$. Given that $\partial_p f(q,t) = 0$, this will not impact the previous arbitrary choice of $\theta_p = 0$. We have

$$\theta_a = \begin{bmatrix} p & 0 & \theta_t \end{bmatrix}. \tag{1.108}$$

At this point, we simply rename the last component to $-H$ and have

$$\theta_a = \begin{bmatrix} p & 0 & -H \end{bmatrix}. \tag{1.109}$$

The form of the vector potential, then, is set by the fact that the flow of S^a is both divergenceless and uniform along the time direction.

This discussion tells us exactly what θ_a is in the case of a single degree of freedom: it is the potential of the displacement field S^a or, equivalently, of the form ω_{ab}. Since we are in the single degree of freedom case, assumption DR is enough to recover Hamiltonian mechanics, and this is the same in the extended-phase-space formulation.

[29] Note that, strictly speaking, while ω_{ab} is a tensor, ϵ_{abc} and S^a are not. Relationship 1.103, then, is valid only if we chose position, momentum and time as variables. Yet, as we will see much later, we can change the position variable and time variable and still have the same expression by redefining momentum and energy appropriately.

We also learned another thing: the Hamiltonian is the time component of the potential, therefore it is not a scalar. That is, if we made a change in the time variable $\hat{t} = \hat{t}(t)$, we would have

$$\hat{H} = -\theta_{\hat{t}} = -d_{\hat{t}}t\,\theta_t = d_{\hat{t}}t\,H. \tag{1.110}$$

It transforms like a scalar only under purely spatial change of variables.

Lagrangian and action

Now that we have seen what θ_a is, we can go back to the Lagrangian and the action. Suppose that we have a path γ, not necessarily an actual evolution, that proceeds along the time variable. That is, we can write

$$\begin{aligned} \gamma &= [q^i(t), p_i(t), t] \\ d\gamma &= d\xi^a = d_t\xi^a\, dt. \end{aligned} \tag{1.111}$$

where we used the time variable as its affine parameter. Note that, in this case, the displacement is along the generic path γ, which is not necessarily an actual evolution. Therefore $d_t\xi^a = S^a$ if and only if γ is an actual evolution of the system. For a generic path we can write:

$$\begin{aligned} L &= p_i v^i - H = \theta_a d_t\xi^a \\ \mathcal{A}[\gamma] &= \int_\gamma L dt = \int_\gamma \theta_a d_t\xi^a dt = \int_\gamma \theta_a d\xi^a = \int_\gamma \theta d\gamma \end{aligned} \tag{1.112}$$

This tells us that

> The action along a path γ is the line integral of the vector potential θ_a of the form ω_{ab}. (1.113)

We now have a precise geometric characterization of the action and of the Lagrangian. As for the physics, it tells us conclusively that both the Lagrangian and the action, by themselves, are unphysical. That is, the value of the Lagrangian at a given position, velocity and time, or the value of the action for a given path is not a physically meaningful value. The vector potential is not a physical quantity: it depends on a choice of gauge, which does not correspond to a physically well-defined object.

We are now in the position to understand why, as we have seen before, the Lagrangian for a given system is not unique: it depends on the vector potential, and the vector potential is not uniquely defined. We can see how the change of Lagrangian 1.94 corresponds to redefining the vector potential as

$$\theta'_a = \theta_a + \partial_a f(q^i, t). \tag{1.114}$$

We have

$$\begin{aligned} L &= \theta_a d_t\xi^a \\ L' &= \theta'_a d_t\xi^a = \theta_a d_t\xi^a + \partial_a f\, d_t\xi^a = L + \partial_{q^i} f\, d_t q^i + \partial_t f\, d_t t \\ &= L + \partial_{x^i} f(x^i, t) v^i + \partial_t f(x^i, t). \end{aligned} \tag{1.115}$$

which is the same expression we had in 1.94.

But if the action and the Lagrangian are unphysical, how can we use them to derive the laws of evolution, which are clearly physical? Note that the laws are not expressed in terms of the action, but in terms of the variation of the action. Let γ', then, be a small variation of the path γ with the same endpoints. Note that γ and γ' form a closed loop. Take a surface Σ that is enclosed by that boundary, we can use Stokes' theorem:

$$\begin{aligned}\delta\mathcal{A}[\gamma] &= \delta\int_\gamma L dt = \delta\int_\gamma \theta_a d\xi^a = \int_\gamma \theta_a d\xi^a - \int_{\gamma'} \theta_a d\xi^a = \oint_{\partial\Sigma} \theta_a d\xi^a = \int_\Sigma \partial_a \wedge \theta_b d\xi^a d\eta^b \\ &= -\int_\Sigma \omega_{ab} d\xi^a d\eta^b,\end{aligned} \tag{1.116}$$

where $d\eta^b$ is the displacement of a point from γ to the corresponding point on the variation γ'. The variation of the action, then, corresponds to the surface integral of ω_{ab}, which is physical. Geometrically, it corresponds to the flow of the evolutions through the surface enclosed by the path and its variation. Note that, because the flow is divergenceless, it does not matter which surface is chosen, since all surfaces that share the same boundaries will correspond to the same flow. This is a striking observation: while the action is unphysical, its variation is physical.

We now are in a perfect position to fully understand the principle of stationary action. This states that actual evolutions are given by paths for which the variation is zero. That is, $\int_\Sigma \omega_{ab} d\xi^a d\eta^b$ has to be zero for all possible $d\eta^b$. This means we must have $\omega_{ab} d\xi^a = 0$. Since S^a is the only degenerate direction for ω_{ab}, we must have $d\xi^a = S^a dt$. In other words, the paths that make the action stationary are exactly the paths whose tangent vector applied to ω_{ab} return zero.

Geometrically, if a path γ is an evolution, it will always be tangent to the displacement field S^a. If we make a small variation γ', the two paths will enclose an infinitesimal strip Σ which will be tangent to S^a as well. Therefore the flow of S^a through Σ will be zero. Given that the flow through Σ is minus the variation of the action, the variation of the action for all evolutions will be zero. Conversely, if a path is not an evolution, at some point its tangent will be different from the displacement field. Therefore we will be able to find a variation γ' for which the surface Σ will "catch" some flow. That is, we will have a variation of the path for which the variation of the action is non-zero. In other words, the action principle is a geometrically roundabout way to ask for those paths that are always tangent to the displacement field S^a, which is divergenceless and has constant flow in time.

Note that the geometric and physical interpretation we have given to the Lagrangian and the action lives in the extended phase space, which is part of the Hamiltonian formulation. In fact, the only place where we need assumption KE is when we want to write $p_i d_t q^i - H$ as a function of velocity instead of momentum. The integral of the vector potential and its variation, in fact, can still be written even in terms of momentum, and therefore the geometric and physical characterization of the principle of stationary action applies for all Hamiltonian systems, not just the Lagrangian subset. In other words, the principle of stationary action is really more a property of Hamiltonian mechanics than Lagrangian mechanics. The only difference is that in Lagrangian mechanics, since assumption KE applies, it can be expressed purely in terms of kinematic variables.

Multiple DOFs

The above discussion captures all the important elements of Lagrangian mechanics and the action principle even if it is limited to the single DOF case. Still, we should generalize to the

multiple DOFs case. Note that the equations we wrote for the Lagrangian and the action are already in generalized form. The only thing that we need to do is derive the expression for the vector potential θ_a in the general case.

If we compare SF-EPS with SF-ND, we find that the non-temporal components of the form are identical. Therefore, we can still understand it as quantifying the number of configurations within each DOF and the degree of independence across DOFs, while adding the idea that the displacement field does not contribute new configurations. Mathematically, the displacement field identifies the only direction in which the form ω_{ab} is degenerate. Let us see what the non-temporal components tell us in terms of the potential. We have:

$$\begin{aligned}
\omega(e^{q^i}, e^{p_j}) &= (-\partial\theta)_{q^i p_j} = -(\partial_{q^i}\theta_{p_j} - \partial_{p_j}\theta_{q^i}) = \delta^i_j \\
\omega(e^{q^i}, e^{q^j}) &= (-\partial\theta)_{q^i q^j} = -(\partial_{q^i}\theta_{q^j} - \partial_{q^j}\theta_{q^i}) = 0 \\
\omega(e^{p_i}, e^{p_j}) &= (-\partial\theta)_{p_i p_j} = -(\partial_{p_i}\theta_{p_j} - \partial_{p_j}\theta_{p_i}) = 0
\end{aligned} \tag{1.117}$$

We can use our gauge freedom to set $\theta_{p_1} = 0$, much in the same way we did before. We now have $\partial_{q^1}\theta_{p_1} = 0$ and, by the first condition, $\partial_{p_1}\theta_{q^1} = 1$. Integrating, we have $\theta_{q^1} = p_1 + g(q^i, p_2, p_3, ..., t)$ where g is an arbitrary function which we can set to zero, since it corresponds to a choice of gauge. Therefore we have:

$$\theta_a = \begin{bmatrix} p_1 & \theta_{q^2} & \cdots & \theta_{q^n} & 0 & \theta_{p_2} & \cdots & \theta_{p_n} & \theta_t \end{bmatrix}. \tag{1.118}$$

Note that the components for the first degree of freedom do not depend on the other degrees of freedom. That is, for all $i > 1$, $\partial_{q^i}\theta_{q^1} = \partial_{p_i}\theta_{q^1} = \partial_{q^i}\theta_{p_1} = \partial_{p_i}\theta_{p_1} = 0$. But by using conditions 1.117, we find that the converse is true as well: the components of all other degrees of freedom do not depend on the first. That is, for all $i > 1$, $\partial_{q^1}\theta_{q^i} = \partial_{p_1}\theta_{q^i} = \partial_{q^1}\theta_{p_i} = \partial_{p_1}\theta_{p_i} = 0$.

We can then use, again, our gauge freedom with a function that does not depend on the first two variables to set $\theta_{p_2} = 0$. And, with the same reasoning, we will be able to set $\theta_{q^2} = p_2$. And then, again, find that the first two degrees of freedom do not depend on the others, etc. This will exhaust all DOFs and we can set $\theta_t = -H$ as before. This will find 1.101.

To recap, we have found that the expression of SF-EPS is equivalent to assuming DR and IND and is also equivalent to 1.101. Note that the expression of ω_{ab} and θ_a are in terms of specific coordinates, while the fact that ω_{ab} is the exterior derivative of θ_a, instead, is coordinate independent. However, there is a characterization of ω_{ab} that is fully coordinate independent.

Since ω_{ab} admits a potential, its exterior derivative is zero. We also have that the displacement field identifies the only degenerate direction. That is, if $v^a\omega_{ab} = 0$ then $v^a = fS^a$ for some scalar function f. Physically, this corresponds to saying that temporal displacement is the only direction that does not provide independent configurations: as we said before, time does not provide new possible configurations for the system. It turns out that these are enough conditions to find canonical coordinates $[q^i, p_i, t]$ such that we can express ω_{ab} as SF-EPS.[30] Therefore

> The two-form ω_{ab} that quantifies independent configurations is closed (i.e. has zero exterior derivative) and its only direction of degeneracy is identified by the displacement field S^a (DI-SYME)

[30] This result is known as Darboux's theorem. Unfortunately we haven't been able to find a proof that is short and/or physically significant, though we suspect such a proof should exist.

is an equivalent characterization of Hamiltonian mechanics in the extended phase space.

The above condition must imply both DR and IND. We can break down the two contributions in the following way. Suppose we have a system with n independent degrees of freedom, meaning the extended phase space is of dimension $N = 2n + 1$. Assumption DR tells us we have a way to measure the flow of the evolutions over a hyper-surface, which corresponds to an $(N-1)$-form $\Omega_{a_1 \cdots a_{2n}}$ such that:

$$S^{a_1} \Omega_{a_1 \cdots a_{2n}} = 0. \tag{1.119}$$

This is just stating that, given that $\Omega_{a_1 \cdots a_{2n}}$ measures the flow through an infinitesimal $2n$-dimensional parallelepiped, if one of the sides is along the direction of flow S^a then the flow will be parallel to the parallelepiped. If the space is charted by position, momentum and time, we can write:

$$\begin{aligned} &\Omega_{a_1 \cdots a_{2n}} = \epsilon_{a_1 \cdots a_{2n} a_{2n+1}} S^{a_{2n+1}} \\ &\int \epsilon_{a_1 \cdots a_{2n} a_{2n+1}} S^{a_{2n+1}} d\xi^{a_1} \cdots d\xi^{a_{2n}} = \int \Omega_{a_1 \cdots a_{2n}} d\xi^{a_1} \cdots d\xi^{a_{2n}}. \end{aligned} \tag{1.120}$$

On the right side, we see that we are really integrating the flow S^a through the hypersurface defined by the differentials $d\xi^{a_i}$, which corresponds to the integral of $\Omega_{a_1 \cdots a_{2n}}$ through the hypersurface. Note that, under DR, we have one and only one evolution for each state, therefore measuring the flow of evolutions is equivalent to measuring the number of states that travel through the surface.

As we said, this setup was motivated by assumption DR alone. If we add IND, then we can write:

$$\Omega_{a_1 \cdots a_{2n}} = \omega_{a_1 a_2} \wedge \omega_{a_3 a_4} \wedge \cdots \wedge \omega_{a_{2n-1} a_{2n}}. \tag{1.121}$$

This tells us that the total state count becomes the product of the configurations along each degree of freedom.

The geometric and physical understanding of the action and the Lagrangian are the same as in the single DOF case: the action is the line integral of the vector potential θ_a and the variation of the action is minus the surface integral of ω_{ab} between the path and its variation. The only difference is that this corresponds not to the flow of total states, but to the flow of configurations for a single degree of freedom. Still, the flow is zero only if the path is everywhere tangent to the displacement field S^a, and therefore if the path is an actual evolution of the system.

1.8 Full kinematic equivalence and massive particles

In the previous section we recovered Lagrangian mechanics by formulating the principle of stationary action in Hamiltonian form, and then used assumption KE in form WKE-INV simply to express it in terms of the kinematic variables. This is the weakest use of the assumption, as it only assumes we have an invertible map between velocity and momentum. However, we will find that this is not enough to define a meaningful count of configurations over kinematic variables, which requires states to be uniformly distributed over velocity. This leads to a stronger version of assumption KE, which requires the map between velocity and conjugate momentum to be linear, which in turn fixes the dynamics to the one of massive particles under scalar and vector potential forces.

Linear map between linear spaces

Condition WKE-INV imposes, at every position, an invertible relationship $p_i = p_i(q^j, v^j)$ between conjugate momentum and velocity which can be arbitrary. Note that both velocity and momentum are linear objects, and therefore it would seem natural to require a linear relationship between the two. Though we do not have, at this point, a clear physical reason to impose this restriction, let's see what it would entail.

If the relationship between momentum and velocity at every point is linear, we have that the Jacobian is only a function of position. Moreover, since it is a linear map between a vector and a covector, the map must be a tensor. We have

$$\partial_{v^i} p_j = m g_{ij} \tag{1.122}$$

where m is a constant that transforms units of velocity to units of momentum while g_{ij} is the actual linear map in terms of spatial coordinates. Since $p_j = \partial_{v^j} L$, we have

$$m g_{ij} = \partial_{v^i} p_j = \partial_{v^i} \partial_{v^j} L = \partial_{v^j} \partial_{v^i} L = \partial_{v^j} p_i = m g_{ji} \tag{1.123}$$

and find that the tensor g_{ij} is symmetric.

If we integrate equation 1.122 we have

$$p_i = m g_{ij} v^j + \mathfrak{q} A_i \tag{1.124}$$

where A_i are arbitrary functions of position and $\mathfrak{q}$ is an arbitrary constant. Since g_{ij} and A_i are functions of position only, we can see that, if we fix position, the relationship is a line where $m g_{ij}$ is the slope and A_i the value of momentum for zero velocity. Note that

$$v^i = d_t q^i = \partial_{p_i} H = \frac{1}{m} g^{ij} (p_j - \mathfrak{q} A_j). \tag{1.125}$$

If we integrate again we find

$$H = \frac{1}{2m} (p_i - \mathfrak{q} A_i) g^{ij} (p_j - \mathfrak{q} A_j) + V \tag{1.126}$$

where V is yet another arbitrary function. This is exactly the Hamiltonian for a massive particle under scalar and vector potential forces: m is the mass, g_{ij} is the metric tensor, $\mathfrak{q}$ is the charge, A_i is a vector potential and V is the scalar potential.[31] Therefore, we saw that imposing a linear relationship between velocity and conjugate momentum gives us massive particles under potential forces.

Conversely, if we start by imposing the above Hamiltonian for massive particles under potential forces, we find a linear relationship between conjugate momentum and velocity. Therefore condition

> There is a linear relationship between conjugate momentum and velocity (FKE-LIN)

and condition

> The system under study is a massive particle under scalar and vector potential forces (FKE-POT)

[31] Note that we have not separated the charge from the potentials.

are equivalent. At this point, having seen enough of reverse physics, it should be clear that this can't simply be a coincidence and it warrants more exploration.

Recall, that in 1.92 we saw that the Jacobian determinant of the transformation from state variables to kinematic variables was equal to the Jacobian determinant of the relationship between velocity and momentum. Since mg_{ij} is the Jacobian between velocity and momentum, we find

$$\begin{aligned} \partial_{p_i} v^j &= \frac{1}{m} g^{ij} \\ |J| = |\partial_{p_i} v^j| &= \frac{|g^{ij}|}{m} = \frac{|g_{ij}|^{-1}}{m} = \frac{1}{m\,|g_{ij}|}. \end{aligned} \tag{1.127}$$

Recall that the existence of a non-singular Jacobian is what allowed us to express all differential objects (e.g. densities, phase-space areas/volumes, the symplectic form) in terms of kinematic variables. This forces those expressions to be in terms of position only. We therefore have that the following conditions are equivalent to each other and equivalent to FKE-LIN and FKE-POT.

The Jacobian of the transformation between state variables and kinematic variables is a non-singular function of position only. (FKE-NSIN)

Densities over phase space can be expressed in terms of position and velocity by rescaling the value at each point: $\rho(x^i, v^j)|J(x^i)| = \rho(q^i, p_j)$. (FKE-DEN)

Areas and volumes in phase space can be expressed in kinematic variables, and the transformation depends on position only: $dx^1 \cdots dx^n\, dv^1 \cdots dv^n = |J(x^i)|\, dq^1 \cdots dq^n\, dp_1 \cdots dp_n$. (FKE-VOL)

The symplectic form ω_{ab} can be expressed in kinematic variables, and its components are a linear function of velocity. (FKE-SYMP)

To understand better these conditions, consider a density ρ_{qp} over phase space and its expression ρ_{xv} over kinematic variables. Since there is a factor of a Jacobian determinant between the two, if the density takes the same value between two states as expressed in position and momentum, the density as defined over position and velocity will not necessarily have the same value. Position and momentum have a special property in that the density of states is uniform in terms of those variables, and therefore comparing densities in those variables is simply a matter of comparing the value of the density. Canonical coordinates are exactly those state variables for which this property holds. While kinematic variables are, in general, not canonical, the linearity between velocity and momentum imposes a lesser version of this property: it makes it so that, at least at the same position, we can compare areas and densities. Therefore if the density for two different values of velocity at the same point matches, $\rho_{xv}(x^i, v_1^j) = \rho_{xv}(x^i, v_2^j)$, then we really have the same density of states. Therefore

Density expressed in velocity at the same position is proportional to the density over states (FKE-PROP)

is another equivalent condition.

In this formulation, one starts to wonder whether the lack of this property would make physical sense. We can understand that non-linear transformations of position, since they stretch and shrink space differently at different points, would make it more complicated to

understand whether two densities at two different points are the same. On the other hand, velocity is defined locally over infinitesimal changes of position, therefore non-linear changes in velocity do not arise when changing units and coordinates.

Mass and inertial frames

Another way of looking at it is the following. Suppose we have a density $\rho_{qp}(q^i, p_j)$ defined over a finite region of phase space that is constant along momentum. That is, $\rho_{qp}(q^i, p_j^1) = \rho_{qp}(q^i, p_j^2)$ for all p_j^1 and p_j^2. Assuming we have a linear relationship between momentum and velocity we have

$$\begin{aligned}\rho_{xv}(x^i, v_1^j) &= m|g_{ij}|\rho_{qp}(q^i, mg_{ij}v_1^j + \mathfrak{q}A_i) = m|g_{ij}|\rho_{qp}(q^i, mg_{ij}v_2^j + \mathfrak{q}A_i) \\ &= \rho_{xv}(x^i, v_2^j).\end{aligned} \tag{1.128}$$

Therefore if we have a linear map between velocity and momentum, uniform distributions along momentum correspond to uniform distributions along velocity. The converse is also true: if uniform distributions along momentum correspond to uniform distributions along velocity, then, the Jacobian of the transformation cannot depend on either, and it must depend on position only. Therefore

> Uniform distributions along momentum correspond to uniform distributions along velocity (FKE-UNIF)

is another equivalent characterization of FKE-LIN. Failure of this condition, then, would mean that physical states are not uniformly distributed over velocity at the same position, and some velocities correspond to higher state densities than others. This does not sound like something that would mesh well with the principle of relativity.

We can go one step further. Suppose we can find a coordinate system for which g_{ij} over a finite region is the identity matrix δ_{ij}. Physically, this corresponds to a Cartesian coordinate system in an inertial frame. In this frame, $|g_{ij}| = 1$ and therefore densities over kinematic variables differ from densities over phase space by a constant of proportionality: the mass. We have the following characterization of mass

> Inertial mass tells us how many states there are per unit of area of position-velocity in an inertial (Cartesian) frame. (IM)

This characterization of mass works even for massless particles, while the standard one does not. In fact, if we define mass to be the resistance of a body to acceleration, we would expect a zero mass body to be extremely easy to accelerate, while this is not the case: a massless particle travels at the same constant speed. On the other hand, as we saw before, massless particles do not satisfy KE, and therefore the state cannot be reconstructed from position and velocity: areas of position-velocity correspond to zero states. So why are more massive bodies harder to accelerate? The more massive the body, the more states per unit of velocity, and therefore reaching the same velocity means going through more states. If we understand force as a change of state per unit time, then we have an intuitive explanation that matches the new characterization.

Note that we focused on Cartesian coordinates, not just inertial frames. That is not a coincidence: the motion for a free particle appears linear and uniform only in these coordinates. That is, trajectories obey the linear expression $x^i = v^i t + x_0^i$ only in Cartesian coordinates.

The first law of Newton, the fact that in an inertial frame a body travels in linear and uniform motion, implicitly implies the ability to use Cartesian coordinates. But, in light of what we have seen before, it also implies the existence of mass. In fact, we can, in true reverse physics spirit, turn this on its head. Suppose that the motion of free particles is linear and uniform. The velocity will remain the same along an evolution; the distance between two particles moving at the same velocity along the same line will also remain the same. Therefore consider the plane charted by position and velocity along a particular direction. Take a parallelogram where two sides are at constant velocity: its area would be the length of those sides Δx times the height, which will be given by the difference in velocity Δv. As time evolves, the sides at constant velocity will move but will remain at the same velocity. The parallelogram will remain a parallelogram with the same height and base: the area is conserved. Now, clearly this is a deterministic and reversible system. Position and velocity identify all states and already provide an area that is conserved over time. It must be, then, that the count of states is proportional to the area in position-velocity, and therefore we have the relationship $p_i = mv^i$. In other words, once we assumed the existence of inertial frames, we already implicitly introduced the idea of inertial mass as characterized above.

The above discussion links the linearity between velocity and momentum to the existence of inertial frames. Therefore

> At each position, there exists a local inertial frame (FKE-INER)

is equivalent to FKE-LIN. Note that here we are claiming that only local inertial frames are needed. To see that, note that a change of position coordinates fully determines the change of both position and velocity. Therefore a change of coordinate can neither introduce nor remove a velocity dependency from the expression $\partial_{v^i} p_j$. If $\partial_{v^i} p_j = m g_{ij}$ has no velocity dependency, we can always find local spatial coordinates such that $g_{ij} = \delta_{ij}$. Therefore, $\partial_{v^i} p_j$ has no velocity dependency if and only if we can find a local inertial (Cartesian) frame.

So we found that FKE-LIN is linked to inertia and the existence of inertial frames, but in what sense is it linked to assumption KE? Suppose one gave us all the kinematics of a system, which are fully defined by the reference frame and all possible trajectories. Velocity, as a quantity, is fully defined in terms of position and time. Now, assumption KE tells us we must be able to recover the full dynamics. We must be able, for example, to convert a distribution over kinematic variables to one over state variables. This is essentially a change of units of the density ρ from position-velocity to position-momentum. But if velocity is a derived quantity, fully defined by position, then it should be the case that the transformation rule of ρ is only a function of position. Therefore condition

> The position fully defines the units of all state variables, and therefore an invertible transformation between momentum and velocity (FKE-UNIT)

is an equivalent characterization of FKE-LIN. Ultimately, this must be the case if KE holds: trajectories in space-time are fully specified by units of position, and since we must be able to reconstruct states from them, so the units of momentum must depend on them.

While this argument works on physics grounds, we cannot carry it out in a mathematically precise way. The issue is that current mathematical structures are devoid of the concept of units and of the type of definitional dependence we used. What does it mean, in a precise way, that velocity is a derived unit from position and time? What does it mean that temperature is

a derived unit from energy and entropy? Covariance and contravariance capture some hint in terms of change of units (i.e. how do units of velocity (or momentum) change if I change units of position?), but this is sufficient only when some initial relationship is assumed. Capturing this initial relationship requires conceptual and mathematical tools that are not currently available.

1.9 Relativistic mechanics

In this section we analyze what happens if we consider change of coordinates that mix space and time. We will see that relativistic elements appear even if a notion of metric tensor is not introduced, and the particle/anti-particle duality emerges. Moreover, under the full kinematic equivalence special relativity is recovered.

Hamiltonian mechanics on the extended phase space

The Hamiltonian formalism is invariant under generic change of spatial variables, but if we want to introduce generic changes of variables that mix space and time we have a problem. In the standard formulation, time is the parameter of the evolution $q^i(t)$, meaning that we can mix the spatial variables q^i while leaving t alone. In the formulation extended by time, time has a double of role of parameter and variable. That is, we write $q^i(t)$ and $t(t)$. Therefore mixing q^i with t will affect the evolution parameter as well. What we need to do is to separate the time variable from the parameter of the evolution. We can group the space-time variables as

$$q^\alpha = [t, q^i]. \tag{1.129}$$

We introduce an affine parameter s, and therefore a trajectory in space-time will be noted as $q^\alpha(s)$. Under a change of coordinate, the variables q^α will mix but the affine parameter s will not change.

Now we have understood how to deal with time, we have to understand how to deal with energy. As we saw before, the Hamiltonian is no longer invariant under transformations that affect time. Moreover, even if we start with a time-independent Hamiltonian, it will not remain so under a generic coordinate transformation that mixes time and space. This means that, during the evolution, energy will not be constant but will need to increase and decrease. Another, more physical, way to look at it is that energy, like momentum, is not an absolute quantity but rather a relative quantity with respect to an observer. If we imagine an observer that is accelerating and decelerating, in the same way that he will see momentum change, he will see the energy change as well. In the same way that momentum is sensitive to changes of units along the corresponding spatial direction, energy is sensitive to change of time units. To be able to characterize these relationships better, we introduce an energy variable E which we group with the momentum variables and, since $-H$ was the time component of the potential θ_a, we write

$$p_\alpha = [-E, p_i] \tag{1.130}$$

so that p_α is a covector.

What we did is add a temporal degree of freedom, and with that addition, the equations look a lot like standard Hamiltonian mechanics for multiple degrees of freedom. We have

$$\begin{aligned} \xi^a &= [q^\alpha \ p_\alpha] \\ S^a &= d_s\xi^a = [d_s q^\alpha \ d_s p_\alpha] \end{aligned} \tag{1.131}$$

We define the potential θ_a[32] and the form ω_{ab}

$$\begin{aligned} \theta_a &= [p_\alpha \ 0] \\ \omega_{ab} &= \partial_a \wedge \theta_b = \begin{bmatrix} \omega_{q^\alpha q^\beta} & \omega_{q^\alpha p_\beta} \\ \omega_{p_\alpha q^\beta} & \omega_{p_\alpha p_\beta} \end{bmatrix} = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}, \end{aligned} \tag{SF-FEPS}$$

and the equations of motion will be

$$S^a\omega_{ab} = \partial_b\mathcal{H}. \tag{1.132}$$

The function $\mathcal{H}$ is called the Hamiltonian constraint. Typically, the constraint is chosen such that

$$\mathcal{H} = 0. \tag{1.133}$$

Relativistic free particle

Given that this formulation is probably unfamiliar to most, let us see how it works for a free particle in a Cartesian (and inertial) reference frame. In this case, it is more convenient to use $q^0 = ct$ as the time variable, which, given 1.110, means using $p_0 = -E/c$ for the energy. The Hamiltonian constraint is

$$\mathcal{H} = \frac{1}{2m}\left(p_\alpha\eta^{\alpha\beta}p_\beta + m^2c^2\right) = \frac{1}{2m}\left(p_i\delta^{ij}p_j - (E/c)^2 + m^2c^2\right), \tag{1.134}$$

where $\eta^{\alpha\beta}$ is the Minkowski metric. Hamilton's equations give

$$\begin{aligned} d_s q^0 &= d_s ct = \partial_{p_0}\mathcal{H} = \frac{1}{2m}\partial_{-E/c}(-(-E/c)^2) = \frac{1}{2m} - 2(-E/c) = \frac{E/c}{m} \\ d_s t &= \frac{E}{mc^2} \\ d_s q^i &= \partial_{p_i}\mathcal{H} = \frac{1}{2m}\partial_{p_i}\left(p_i\delta^{ij}p_j\right) = \frac{p^i}{m} \\ d_s p_0 &= d_s(-E/c) = -\partial_{q^0}\mathcal{H} = 0 \\ d_s p_i &= -\partial_{q^i}\mathcal{H} = 0. \end{aligned} \tag{1.135}$$

Therefore we have

$$p^\alpha = m d_s q^\alpha = m u^\alpha. \tag{1.136}$$

[32] Recall that when phase space was extended with just time, we had $\theta_a = [\theta_{q^i} \ \theta_{p_i} \ \theta_t] = [p_i \ 0 \ -H]$. Now that it is also extended with energy, we have $\theta_a = [\theta_t \ \theta_{q^i} \ \theta_{-E} \ \theta_{p_i}] = [-E \ p_i \ 0 \ 0] = [\theta_{q^\alpha} \ \theta_{p_\alpha}] = [p_\alpha \ 0]$.

We recognize u^α as the four-velocity, and the affine parameter s as proper time. With this in mind, we can rewrite the Hamiltonian constraint as

$$\mathcal{H} = \frac{1}{2m}\left(mu^\alpha \eta_{\alpha\beta} mu^\beta + m^2c^2\right) = \frac{1}{2}m|u|^2 + \frac{1}{2}mc^2, \tag{1.137}$$

which looks like a kinetic energy term constructed from the four-velocity. Setting $\mathcal{H}$ to zero, then, means setting the norm squared of the four-velocity to $-c^2$, which is consistent with special relativity. It also means that

$$\begin{aligned}(-E/c)^2 &= m^2c^2 + p_i p^i \\ E &= \pm\sqrt{c^2|p_i|^2 + (mc^2)^2}.\end{aligned} \tag{1.138}$$

That is, the Hamiltonian constraint sets the relationship between energy and momentum.

The geometry of extended phase space

Note that while the mathematical structure of the extended phase space looks the same at first glance as the one of standard phase space, it has important differences. Let's concentrate on the Hamiltonian constraint. In standard Hamiltonian mechanics, the Hamiltonian H gives a value of energy that is conserved by the system during the evolution, but the system can be in different states with different energy. The Hamiltonian constraint $\mathcal{H}$, instead, is a quantity that is not only conserved by the system during the evolution, it is always the same for the given system. In the example above, the Hamiltonian constraint essentially is a constraint on the mass of the system. Therefore the space of states is not really the full $2n+2$ dimensional manifold charted by time, position, energy and momentum, but it is the $2n+1$ dimensional sub-manifold that is given by the constraint $\mathcal{H} = 0$. The constraint lowers the dimensionality by imposing a relationship between energy and the other state variables. The Hamiltonian constraint, then, plays a double role: as a generator of the evolution over the affine parameter s and as an equation of state of the system.[33]

Since over valid states we will have both $\mathcal{H} = 0$ and $E = H$, we can write

$$\mathcal{H} = \lambda(H - E). \tag{1.139}$$

where λ is a function of the extended phase space. This expression provides a link between the Hamiltonian constraint and the standard Hamiltonian. To understand what λ is, note that

$$d_s t = \partial_{-E}\mathcal{H} = \lambda. \tag{1.140}$$

Therefore λ is the rate of change between time and the affine parameter. Additionally, given that time must still be a possible parameter for the evolution, $t(s)$ must be an invertible strictly monotonic function. Therefore we must always have $d_s t \neq 0$ at least over the region

[33] Note that to apply the constraint to a distribution ρ, one can simply write $\mathcal{H}\rho = 0$. This means that ρ can be different from zero only over the region where $\mathcal{H}$ is zero, and therefore ρ is non-zero only for those points that satisfy the Hamiltonian constraint. Moreover, note that in the case of a free particle in a Cartesian frame, the Hamiltonian constraint $\mathcal{H}\rho = 0$ is the classical analogue of the Klein-Gordon equation $\left(\frac{\hbar^2}{c^2}\partial_t^2 - \hbar^2\nabla^2 + m^2c^2\right)\psi = 0$.

where $E = H$. We can now verify that the new formulation recovers the old.

$$\begin{aligned}
\mathcal{H} &= \lambda(H - E) = 0 \\
E &= H \\
d_t t &= 1 = d_t s \, d_s t = d_t s \, \partial_{-E} \mathcal{H} = d_t s \, \lambda \\
d_t s &= \frac{1}{\lambda} \\
d_t q^i &= d_t s \, d_s q^i = d_t s \, \partial_{p_i} \mathcal{H} = \frac{1}{\lambda} \left(\partial_{p_i} \lambda (H - E) + \lambda \partial_{p_i} H \right) = \partial_{p_i} H \\
d_t p_i &= d_t s \, d_s p_i = -d_t s \, \partial_{q^i} \mathcal{H} = -\frac{1}{\lambda} \left(\partial_{q^i} \lambda (H - E) + \lambda \partial_{q^i} H \right) = -\partial_{q^i} H \\
d_t E &= d_t s \, d_s E = d_t s \, \partial_t \mathcal{H} = \frac{1}{\lambda} \left(\partial_t \lambda (H - E) + \lambda \partial_t H \right) = \partial_t H
\end{aligned} \tag{1.141}$$

Therefore if we set $\mathcal{H} = H - E$, the affine parameter will have to be equal to time up to an arbitrary additive constant, and therefore the formulations are equivalent.

The expression above seems to imply that the Hamiltonian constraint always has to correspond to one Hamiltonian. This is not the case. Suppose we have

$$\mathcal{H} = (H_1 - E)(H_2 - E) \tag{1.142}$$

so that $H_1(\xi^a) \neq H_2(\xi^a)$ at all points. This means that there are two regions of the extended phase space that satisfy the Hamiltonian constraints, one for H_1 and one for H_2. These two regions are disconnected, therefore states from different regions cannot be connected by Hamiltonian evolution. They essentially represent two different types of system encoded into the same equation. In principle, the same idea can be extended to have more than two regions.

While this may seem just a mathematical artifact, note that this is exactly what happens for the Hamiltonian constraint of a free particle. In fact we can write:

$$\mathcal{H} = \frac{1}{2mc^2} \left(\sqrt{c^2 |p_i|^2 + (mc^2)^2} + E \right) \left(\sqrt{c^2 |p_i|^2 + (mc^2)^2} - E \right) \tag{1.143}$$

Given that the solutions for energy have opposite sign, one finds

$$\lambda = \frac{1}{2mc^2} (E + E) = \frac{E}{mc^2} \tag{1.144}$$

in agreement with what we found before.

Note that for a free particle, similarly to what one finds in quantum field theory, we have both positive and negative energy solutions. Given that $d_s t = E/mc^2$, what happens is the affine parameter is anti-aligned with respect to time. For the negative energy solutions, then, s will be minus proper time instead of proper time.[34]

In general, $d_s t$ may be positive or negative, which corresponds to the affine parameter being aligned or anti-aligned with respect to time. Note that since $d_s t$ cannot be equal to zero, therefore states for which $d_s t > 0$ can never be connected with states for which $d_s t < 0$.

[34] This is a much more precise version of the claim that anti-particles "travel backwards in time". They do not. What happens is that the function $t(s)$ is parameterized in the opposite direction. However, the affine parameter s is not physically significant.

Therefore these two types of states must always exist in different regions of the extended phase space. In analogy with what happens in quantum field theory, we call particle states those for which $d_s t > 0$ and anti-particle states those for which $d_s t < 0$. We therefore find the following

Insight 1.145. *A frame-invariant notion of determinism and reversibility (i.e. allowing generalized coordinate transformations in Hamiltonian mechanics) gives us the notion of anti-particles, even in classical particle mechanics.*

Now that we got a better feel for what the Hamiltonian constraint is and how it works, let's go back to explore the geometry and physical significance of the extended phase space. Looking at ω_{ab} it may seem that adding the temporal degree of freedom means just adding another independent DOF, but that is not the case. If we start with standard phase space and add time, we are not adding new independent configurations: if we have deterministic and reversible motion, given the state at one time, we know the state at all times. So, while adding time allows us to talk about all states at all times, we are not really adding new states because states are defined and counted at equal time. In phase space extended by time only, this was captured by the fact that the form ω_{ab} applied to the displacement field always gives zero. If we now add energy, the conjugate of time, things are even more constrained. The energy, in fact, is a function of the state of the system at a particular time. The addition of energy, then, adds no configurations at all.

Now that it is clear that $(t, -E)$ do not constitute an independent degree of freedom, we should understand the significance of the minus sign. As we saw, it was needed so that p_α is a covector, but that does not tell us anything in terms of the geometry. What we are really saying is that an infinitesimal rectangle of size dt and dE has a negative contribution to the count of states. Shouldn't it have no contribution at all?

We saw that orthogonality in phase space means DOF independence. If we have two independent DOFs we can write:

$$\omega(dq^1 + dq^2, dp_1 + dp_2) = dq^1 dp_1 + dq^2 dp_2. \tag{1.146}$$

The above expression, then, can be understood as an areal version of Pythagoras's theorem: instead of summing the square distances, we are summing areas directly. That is, given the orthogonality of the independent DOFs, we have a right triangle-like structure where the two sides are the areas in each DOF and the hypotenuse is the area determined by ω. Now, because spatial and temporal degrees of freedom are not independent, i.e. they are not orthogonal, the form of the above expression cannot be the correct one for the temporal DOF.

The naive consideration would simply be to disregard the temporal degree of freedom, since time does not contribute new states, and set

$$\omega(dq + dt, dp + dE) = dq\, dp. \tag{1.147}$$

However, this does not work either. This would say that the temporal DOF does not identify states at all, which is not the case. Suppose we study a free particle with one degree of freedom. Let's assume that for $t = 0$, $q = 0$. We have

$$\begin{aligned} q &= \frac{p}{m} t \\ E &= \frac{p^2}{2m}. \end{aligned} \tag{1.148}$$

If we look at the region where momentum is positive, the relationship is bijective. Therefore time and energy can be used to identify, and therefore count, states. The issue is that those states are not new states, they are the same ones that are identified, and counted, by position and momentum.

If we look at the expressions derived before, we have

$$\omega(dq + dt, dp + dE) = dq\,dp + dt\,d(-E) = dq\,dp - dt\,dE. \tag{1.149}$$

We can rearrange this expression as

$$dq\,dp = \omega(dq + dt, dp + dE) + dt\,dE. \tag{1.150}$$

Now, compare 1.146 with 1.150. In the second expression, the area in the spatial DOF is the hypotenuse, while the area on the temporal DOF and the count of states given by ω are the sides. That is, while the temporal DOF and the spatial DOF are not orthogonal, the temporal DOF is orthogonal to the surface where states are counted. If we have a region defined at equal time, for example, the count of states reduces to the count of spatial configurations, which makes sense. In the general case, the differentials dt and dE are defined over surfaces at constant (q, p), while the differentials dq and dp are defined over surfaces at constant (t, E). These, as we said, are not orthogonal. This means that the area identified by dq and dp may have a non-zero projection over the temporal degree of freedom. Given that the state count should be defined at equal time, the area given by $dq\,dp$ does not always properly count states. If we want the count of states to be defined at equal time, this needs to be done on the surface that is orthogonal to the surface where dt and dE are defined, which therefore forms a right triangle-like structure with the spatial and temporal degrees of freedom.

The minus sign for the energy, then, has both a geometrical and physical meaning. If we used E instead of $-E$ as the variable, the minus sign would show up in the form ω_{ab}, which would probably be better as it would make the geometrical feature more clear. However, $p_\alpha = [E\ p_i]$ would not match the definition used in relativity and it would not form a covector, so it would be confusing in a different way. If there is a better overall notation and grouping, we haven't yet found it. At any rate, we have found that

Insight 1.151. *A frame-invariant notion of determinism and reversibility (i.e. allowing generalized coordinate transformations in Hamiltonian mechanics) gives elements of special relativity (i.e. energy-momentum four-vector), even without the notion of a metric tensor.*

What is interesting is that we didn't need to add any further assumption: DR and IND are sufficient. The only thing we needed to add was the ability to make generalized space-time transformations.

Relativistic kinematics

It is time to add assumption KE. In this generalized setting, the trajectory, and therefore the velocity, will be in terms of the affine parameter s. Therefore we set

$$d_s q^\alpha = d_s t\, d_t q^\alpha = d_s t\,[d_t q^0, d_t q^i] = d_s t\,[d_t q^0, v^i]. \tag{1.152}$$

Given that the Hamiltonian constraint on the extended phase space plays the same role as the Hamiltonian on the standard phase space, we will have an equivalent principle of stationary

action once KE is taken. The action can be written as

$$A[\gamma] = \int_\gamma L ds = \int_\gamma (p_\alpha u^\alpha - \mathcal{H})\, ds \tag{1.153}$$

Under the full kinematic assumption, we have

$$\begin{aligned} \partial_{u^\alpha} p_\beta &= m g_{\alpha\beta} \\ p_\alpha &= m g_{\alpha\beta} u^\beta + \mathfrak{q} A_\alpha \\ u^\alpha &= d_s q^\alpha = \partial_{p_\alpha} \mathcal{H} = \frac{1}{m} g^{\alpha\beta}(p_\beta - \mathfrak{q} A_\beta) \\ \mathcal{H} &= \frac{1}{2m}(p_\alpha - \mathfrak{q} A_\alpha)\, g^{\alpha\beta}\left(p_\beta - \mathfrak{q} A_\beta\right) + U \end{aligned} \tag{1.154}$$

If we set $U = \frac{1}{2}mc^2$, this is the Hamiltonian constraint for a massive particle under potential forces.[35]

As we saw, the best way to understand the geometry of phase space is to use the state variables q^α and p_α. However, understanding the physics is tricky as conjugate momentum is a gauge dependent quantity. One thing we can do is actually work on phase space with kinematic quantities, and express elements like θ_a and ω_{ab} in those variables. We have:

$$\begin{aligned} q^\alpha &= x^\alpha \\ p_\alpha &= m g_{\alpha\beta} u^\beta + \mathfrak{q} A_\alpha \\ x^\alpha &= q^\alpha \\ u^\beta &= \frac{1}{m} g^{\beta\alpha}(p_\alpha - \mathfrak{q} A_\alpha) \\ \mathcal{H} &= \frac{1}{2} m u^\alpha g_{\alpha\beta} u^\beta + \frac{1}{2} mc^2 \end{aligned} \tag{1.155}$$

Note how the Hamiltonian constraint, in kinematic variables, is always the same, regardless of the forces acting on the particle.

To calculate the expressions of θ_a and ω_{ab} in kinematic variables, we first see how the covector basis transforms. We have

$$\begin{aligned} e^{q^\alpha} &= \partial_{x^\beta} q^\alpha e^{x^\beta} + \partial_{u^\gamma} q^\alpha e^{u^\gamma} = \delta^\alpha_\beta e^{x^\beta} + 0 e^{u^\gamma} = e^{x^\alpha} \\ e^{p_\alpha} &= \partial_{x^\beta} p_\alpha e^{x^\beta} + \partial_{u^\gamma} p_\alpha e^{u^\gamma} \\ &= \left(m \partial_{x^\beta} g_{\alpha\gamma} u^\gamma + \mathfrak{q} \partial_{x^\beta} A_\alpha\right) e^{x^\beta} + m g_{\alpha\gamma} e^{u^\gamma} \end{aligned} \tag{1.156}$$

Now we can express the forms in terms of state variables and perform variable substitution.

$$\theta = \theta_a e^a = p_\alpha e^{q^\alpha} = (m g_{\alpha\beta} u^\beta + \mathfrak{q} A_\alpha) e^{x^\alpha}$$

$$\begin{aligned} \omega &= \omega_{ab} e^a \otimes e^b = \omega_{q^\alpha p_\beta} e^{q^\alpha} \otimes e^{p_\beta} + \omega_{p_\alpha q^\beta} e^{p_\alpha} \otimes e^{q^\beta} = e^{q^\alpha} \otimes e^{p_\alpha} - e^{p_\alpha} \otimes e^{q^\alpha} \\ &= e^{x^\alpha} \otimes \left[\left(m \partial_{x^\beta} g_{\alpha\gamma} u^\gamma + \mathfrak{q} \partial_{x^\beta} A_\alpha\right) e^{x^\beta} + m g_{\alpha\gamma} e^{u^\gamma}\right] - \left[\left(m \partial_{x^\beta} g_{\alpha\gamma} u^\gamma + \mathfrak{q} \partial_{x^\beta} A_\alpha\right) e^{x^\beta} + m g_{\alpha\gamma} e^{u^\gamma}\right] \otimes e^{x^\alpha} \\ &= \left(m \partial_{x^\beta} g_{\alpha\gamma} u^\gamma + \mathfrak{q} \partial_{x^\beta} A_\alpha\right) e^{x^\alpha} \otimes e^{x^\beta} + m g_{\alpha\beta} e^{x^\alpha} \otimes e^{u^\beta} - \left(m \partial_{x^\beta} g_{\alpha\gamma} u^\gamma + \mathfrak{q} \partial_{x^\beta} A_\alpha\right) e^{x^\beta} \otimes e^{x^\alpha} - m g_{\alpha\beta} e^{u^\beta} \otimes e^{x^\alpha} \\ &= \left(m u^\gamma \left(\partial_{x^\beta} g_{\alpha\gamma} - \partial_{x^\alpha} g_{\beta\gamma}\right) + \mathfrak{q}\left(\partial_{x^\beta} A_\alpha - \partial_{x^\alpha} A_\beta\right)\right) e^{x^\alpha} \otimes e^{x^\beta} + m g_{\alpha\gamma} e^{x^\alpha} \otimes e^{u^\gamma} - m g_{\alpha\gamma} e^{u^\gamma} \otimes e^{x^\alpha} \end{aligned}$$

[35] An open problem is understanding whether U is constrained to be a constant or not.

(1.157)

We introduce the following two expressions

$$\begin{aligned} F_{\alpha\beta} &= \partial_{x^\alpha} A_\beta - \partial_{x^\beta} A_\alpha \\ G_{\alpha\beta\gamma} &= \partial_{x^\alpha} g_{\beta\gamma} - \partial_{x^\beta} g_{\alpha\gamma} \end{aligned} \tag{1.158}$$

We have

$$\begin{aligned} \theta_a &= [m g_{\alpha\beta} u^\beta + \mathfrak{q} A_\alpha \ 0] \\ \omega_{ab} &= \begin{bmatrix} -m G_{\alpha\beta\gamma} u^\gamma - \mathfrak{q} F_{\alpha\beta} & m g_{\alpha\beta} \\ -m g_{\alpha\beta} & 0 \end{bmatrix} \end{aligned} \tag{1.159}$$

We recognize $F_{\alpha\beta}$ as the electromagnetic field tensor. However, it is still an open problem what $G_{\alpha\beta\gamma}$ represents. It has a direct relationship with the Christoffel symbols $\Gamma_{\alpha\beta\gamma}$ [36]

$$\begin{aligned} G_{\alpha\beta\gamma} &= \partial_\alpha g_{\beta\gamma} - \partial_\beta g_{\alpha\gamma} = \frac{1}{2}(\partial_\alpha g_{\beta\gamma} - \partial_\beta g_{\alpha\gamma}) - \frac{1}{2}(\partial_\beta g_{\alpha\gamma} - \partial_\alpha g_{\beta\gamma}) \\ &= \frac{1}{2}(\partial_\gamma g_{\alpha\beta} + \partial_\alpha g_{\beta\gamma} - \partial_\beta g_{\alpha\gamma}) - \frac{1}{2}(\partial_\gamma g_{\alpha\beta} + \partial_\beta g_{\alpha\gamma} - \partial_\alpha g_{\beta\gamma}) \\ &= \Gamma_{\beta\alpha\gamma} - \Gamma_{\alpha\beta\gamma}. \end{aligned} \tag{1.160}$$

If we express the two tensors in covariant components, we find[37]

$$\begin{aligned} G^{\alpha\beta\gamma} &= g^{\alpha\delta} g^{\beta\epsilon} g^{\gamma\zeta} G_{\delta\epsilon\zeta} = g^{\alpha\delta} g^{\beta\epsilon} g^{\gamma\zeta} \partial_\delta g_{\epsilon\zeta} - g^{\alpha\delta} g^{\beta\epsilon} g^{\gamma\zeta} \partial_\epsilon g_{\delta\zeta} \\ &= -g^{\alpha\delta} \partial_\delta g^{\beta\gamma} + g^{\beta\epsilon} \partial_\epsilon g^{\alpha\gamma} \\ &= \partial^\beta g^{\alpha\gamma} - \partial^\alpha g^{\beta\gamma} \end{aligned} \tag{1.161}$$

$$\begin{aligned} F^{\alpha\beta} &= g^{\alpha\gamma} g^{\beta\delta} F_{\gamma\delta} = g^{\alpha\gamma} g^{\beta\delta} \partial_\gamma A_\delta - g^{\alpha\gamma} g^{\beta\delta} \partial_\delta A_\gamma \\ &= g^{\alpha\gamma} \partial_\gamma (g^{\beta\delta} A_\delta) - g^{\beta\delta} \partial_\delta (g^{\alpha\gamma} A_\gamma) - g^{\alpha\gamma} \partial_\gamma g^{\beta\delta} A_\delta + g^{\beta\delta} \partial_\delta g^{\alpha\gamma} A_\gamma \\ &= \partial^\alpha A^\beta - \partial^\beta A^\alpha + \partial^\beta g^{\alpha\gamma} A_\gamma - \partial^\alpha g^{\beta\delta} A_\delta \\ &= \partial^\alpha A^\beta - \partial^\beta A^\alpha + G^{\alpha\beta\gamma} A_\gamma \end{aligned} \tag{1.162}$$

The link between between $F^{\alpha\beta}$ and $G^{\alpha\beta\gamma}$ is not specific to the electromagnetic field. In fact

$$\begin{aligned} \nabla^\alpha A^\beta - \nabla^\beta A^\alpha &= g^{\alpha\gamma} \nabla_\gamma A^\beta - g^{\beta\gamma} \nabla_\gamma A^\alpha \\ &= g^{\alpha\gamma} (\partial_\gamma A^\beta + \Gamma^\beta_{\gamma\delta} A^\delta) - g^{\beta\gamma} (\partial_\gamma A^\alpha + \Gamma^\alpha_{\gamma\delta} A^\delta) \\ &= \partial^\alpha A^\beta - \partial^\beta A^\alpha + g^{\alpha\delta} g^{\beta\epsilon} A^\gamma (\Gamma_{\epsilon\delta\gamma} - \Gamma_{\delta\epsilon\gamma}) \\ &= \partial^\alpha A^\beta - \partial^\beta A^\alpha + G^{\alpha\beta\gamma} A_\gamma \end{aligned} \tag{1.163}$$

[36] It is not the torsion as it anti-symmetrizes the first two indexes of the Christoffel symbols, while the torsion uses the second two. In fact, we used the expression for a connection with no torsion to derive the expression.

[37] The derivation uses the following relationship

$$\partial_\alpha g^{\beta\gamma} = -g^{\beta\delta} g^{\gamma\epsilon} \partial_\alpha g_{\delta\epsilon}$$

which can be derived from the following

$$0 = \partial_\alpha \delta^\beta_\epsilon = \partial_\alpha (g^{\beta\delta} g_{\delta\epsilon}) = g_{\delta\epsilon} \partial_\alpha g^{\beta\delta} + g^{\beta\delta} \partial_\alpha g_{\delta\epsilon} = g^{\gamma\epsilon} g_{\delta\epsilon} \partial_\alpha g^{\beta\delta} + g^{\gamma\epsilon} g^{\beta\delta} \partial_\alpha g_{\delta\epsilon} = \partial_\alpha g^{\beta\gamma} + g^{\gamma\epsilon} g^{\beta\delta} \partial_\alpha g_{\delta\epsilon}$$

The expression $G_{\alpha\beta\gamma}$ is linked to a sort of covariant exterior derivative for vectors. Unfortunately, we do not truly understand its geometrical significance.

To find the equations for motions, we first calculate the Poisson brackets between the kinematic variables.

$$\{x^\alpha, x^\beta\} = \{q^\alpha, q^\beta\} = 0 \tag{1.164}$$

$$\begin{aligned}\{x^\alpha, u^\beta\} &= \{q^\alpha, \frac{1}{m}g^{\beta\gamma}(p_\gamma - \mathfrak{q}A_\gamma)\} = \{q^\alpha, \frac{1}{m}g^{\beta\gamma}p_\gamma\} - \{q^\alpha, \frac{\mathfrak{q}}{m}g^{\beta\gamma}A_\gamma\} \\ &= \frac{1}{m}g^{\alpha\beta}\end{aligned} \tag{1.165}$$

$$\begin{aligned}\{u^\alpha, u^\beta\} &= \frac{1}{m^2}\{g^{\alpha\gamma}p_\gamma - \mathfrak{q}A^\alpha, g^{\beta\delta}p_\delta - \mathfrak{q}A^\beta\} \\ &= \frac{1}{m^2}\left[\{g^{\alpha\gamma}p_\gamma, g^{\beta\delta}p_\delta\} - \{g^{\alpha\gamma}p_\gamma, \mathfrak{q}A^\beta\} - \{\mathfrak{q}A^\alpha, g^{\beta\delta}p_\delta\} + \{\mathfrak{q}A^\alpha, \mathfrak{q}A^\beta\}\right] \\ &= \frac{1}{m^2}\left[g^{\alpha\gamma}\{p_\gamma, p_\delta\}g^{\beta\delta} + p_\gamma\{g^{\alpha\gamma}, p_\delta\}g^{\beta\delta} + g^{\alpha\gamma}\{p_\gamma, g^{\beta\delta}\}p_\delta + p_\gamma\{g^{\alpha\gamma}, g^{\beta\delta}\}p_\delta\right. \\ &\left.-g^{\alpha\gamma}\{p_\gamma, \mathfrak{q}A^\beta\} - p_\gamma\{g^{\alpha\gamma}, \mathfrak{q}A^\beta\} - \{\mathfrak{q}A^\alpha, p_\delta\}g^{\beta\delta} - \{\mathfrak{q}A^\alpha, g^{\beta\delta}\}p_\delta\right] \\ &= \frac{1}{m^2}\left[G^{\alpha\beta\gamma}p_\gamma + \mathfrak{q}(\partial^\alpha A^\beta - \partial^\beta A^\alpha)\right] \\ &= \frac{1}{m^2}\left[G^{\alpha\beta\gamma}m(g_{\gamma\delta}u^\delta + \mathfrak{q}A_\gamma) + \mathfrak{q}(\partial^\alpha A^\beta - \partial^\beta A^\alpha)\right] \\ &= \frac{1}{m^2}\left[G^{\alpha\beta\gamma}mg_{\gamma\delta}u^\delta + \mathfrak{q}(G^{\alpha\beta\gamma}A_\gamma + \partial^\alpha A^\beta - \partial^\beta A^\alpha)\right] \\ &= \frac{1}{m^2}\left[G^{\alpha\beta\gamma}mg_{\gamma\delta}u^\delta + \mathfrak{q}F^{\alpha\beta}\right]\end{aligned} \tag{1.166}$$

We find the evolution of the kinematic variables by calculating the Poisson bracket with the Hamiltonian constraint.

$$\begin{aligned}d_s x^\alpha = \{x^\alpha, \mathcal{H}\} &= \frac{1}{2}m\{x^\alpha, u^\beta g_{\beta\gamma}u^\gamma\} = mu^\beta g_{\beta\gamma}\{x^\alpha, u^\gamma\} \\ &= mu^\beta g_{\beta\gamma}\frac{1}{m}g^{\alpha\gamma} = u^\alpha\end{aligned} \tag{1.167}$$

$$\begin{aligned}d_s u^\alpha = \{u^\alpha, \mathcal{H}\} &= \frac{1}{2}m\{u^\alpha, u^\beta g_{\beta\gamma}u^\gamma\} \\ &= mu^\beta g_{\beta\gamma}\{u^\alpha, u^\gamma\} + \frac{1}{2}mu^\beta u^\gamma\{u^\alpha, g_{\beta\gamma}\} \\ &= mu^\beta g_{\beta\gamma}\frac{1}{m^2}(G^{\alpha\gamma\delta}mg_{\delta\epsilon}u^\epsilon + \mathfrak{q}F^{\alpha\gamma}) - \frac{1}{2}mu^\beta u^\gamma\frac{1}{m}g^{\alpha\delta}\partial_\delta g_{\beta\gamma} \\ &= u^\beta u^\epsilon g_{\beta\gamma}g_{\epsilon\delta}G^{\alpha\gamma\delta} - \frac{1}{2}u^\beta u^\gamma g^{\alpha\delta}\partial_\delta g_{\beta\gamma} + \frac{q}{m}F^{\alpha\gamma}g_{\gamma\beta}u^\beta \\ &= u^\beta u^\epsilon g^{\alpha\gamma}G_{\gamma\beta\epsilon} - \frac{1}{2}u^\beta u^\gamma g^{\alpha\delta}(\partial_\beta g_{\gamma\delta} - \partial_\beta g_{\gamma\delta} + \partial_\delta g_{\beta\gamma}) + \frac{q}{m}F^{\alpha\gamma}g_{\gamma\beta}u^\beta\end{aligned}$$

$$
\begin{aligned}
&= u^\beta u^\gamma g^{\alpha\delta} G_{\delta\beta\gamma} - u^\beta u^\gamma g^{\alpha\delta} \frac{1}{2}(\partial_\gamma g_{\delta\beta} + \partial_\delta g_{\beta\gamma} - \partial_\beta g_{\gamma\delta}) + \frac{q}{m} F^{\alpha\gamma} g_{\gamma\beta} u^\beta \\
&= u^\beta u^\gamma g^{\alpha\delta} (\Gamma_{\beta\delta\gamma} - \Gamma_{\delta\beta\gamma} - \Gamma_{\beta\delta\gamma}) + \frac{q}{m} F^{\alpha\gamma} g_{\gamma\beta} u^\beta \\
&= -u^\beta u^\gamma g^{\alpha\delta} \Gamma_{\delta\beta\gamma} + \frac{q}{m} F^{\alpha\gamma} g_{\gamma\beta} u^\beta
\end{aligned}
$$

$$
D_s u^\alpha = d_s u^\alpha + u^\beta d_s x^\gamma \Gamma^\alpha_{\beta\gamma} = \frac{q}{m} F^{\alpha\gamma} g_{\gamma\beta} u^\beta \tag{1.168}
$$

These are geodesic equations modified by the electromagnetic force, which are consistent with general relativity.

We have found that the addition of assumption KE in the generalized case gives us relativistic mechanics and only relativistic mechanics. There was no choice at any point, therefore, in this sense, relativistic mechanics is the only option that works.

Insight 1.169. *Relativistic mechanics is a consequence of DR, IND and KE.*

Let's see how the Minkowski metric and the speed of light emerge from what we have already discussed. The strong kinematic assumption imposes a linear relationship between velocity and momentum which, in the absence of forces, can be written as

$$
p_\alpha = m g_{\alpha\beta} u^\alpha. \tag{1.170}
$$

Given that $g_{\alpha\beta}$ is a symmetric tensor that depends only on space-time, it can be diagonalized at a point P with a suitable coordinate choice. While the velocity is in terms of an affine parameter s, we can express it in terms of time since $d_s x^\alpha = d_s t \, d_t x^\alpha = \lambda d_t x^\alpha$. Since we can set $x^0 = t$, the above equation becomes

$$
\begin{aligned}
-E &= m \, g_{00} \, \lambda \, d_t t = \lambda m \, g_{00} \\
p_i &= m \, g_{ii} \, \lambda \, d_t x^i = \lambda m \, g_{ii} \, v^i.
\end{aligned} \tag{1.171}
$$

Dimensional analysis on the spatial components tells us that g_{ii} must be pure numbers, under the assumption that s has dimensions of time. For the time component, instead, dimensional analysis tells us that g_{00} must be the square of a velocity. As we saw before, solution with positive energy are those for which s and λ are aligned, therefore g_{00} must be a negative quantity. Therefore we set $g_{00} = -c^2$, and we recognize c as the speed of light. Note that, technically, the constant does not play at all a role of speed in this discussion, just as a conversion factor from the temporal component of the four-velocity to the energy.

We can change units so that $g_{00} = -1$ and $g_{ii} = 1$. For the spatial components, it is just a matter of rescaling the units. For time, units need to be changed as well by setting $p_0 = -E/c$ and $x^0 = ct$. Note that the product of the two remains of the same unit, and the two are still conjugate as space variables. Therefore we have shown that, at every point, we can find a set of space-time coordinates such as $g_{\alpha\beta} = \eta_{\alpha\beta}$ where $\eta_{\alpha\beta}$ is the Minkowki metric. If we use proper time as the affine parameter s, then λ becomes $\gamma = \frac{1}{\sqrt{1+\left(\frac{v^2}{c^2}\right)}}$

The failure of Galilean relativity

Given that our assumptions led directly to relativistic mechanics, non-relativistic mechanics must fail to satisfy the strong kinematic assumption. We have not identified the exact reason of the failure, though we have identified an interesting issue.

Galilean space-time transformations correspond to

$$\begin{aligned} \hat{t} &= t \\ \hat{x} &= x + v_0 t \\ \hat{v} &= v + v_0 \end{aligned} \tag{1.172}$$

The canonical transformation induced by that space-time variable change is the following

$$\begin{aligned} t &= \hat{t} \\ q &= \hat{q} - v_0 \hat{t} \\ \hat{p} &= \partial_{\hat{q}} q\, p + \partial_{\hat{q}} t(-E) = p \\ -\hat{E} &= \partial_{\hat{t}} q\, p + \partial_{\hat{t}} t(-E) = -v_0 p - E \end{aligned} \tag{1.173}$$

which gives us different rules for how momentum and energy transform. Note, for example, that the momentum remains unchanged. The expected rules would be

$$\begin{aligned} \hat{p} &= m\hat{v} = mv + mv_0 = p + mv_0 \\ \hat{E} &= \frac{1}{2} m\hat{v}^2 = \frac{1}{2} m(v + v_0)^2 = \frac{1}{2} mv^2 + mvv_0 + \frac{1}{2} m(v_0)^2 \\ &= E + v_0 p + \frac{1}{2} m(v_0)^2. \end{aligned} \tag{1.174}$$

The difference between the two expressions is a constant, therefore Galilean change of variables together with non-relativistic expressions for momentum and energy are still canonical transformations. However, accommodating that constant requires a change of gauge. That is, if we start in an inertial frame where

$$p_\alpha = m\lambda g_{\alpha\beta} u^\beta \tag{1.175}$$

we end up in another inertial frame where

$$\begin{aligned} p_\alpha &= m\lambda g_{\alpha\beta} u^\beta + A_\alpha \\ A_\alpha &= \left[\frac{1}{2} m|v_0|^2 \;\; mv_0^i \right]. \end{aligned} \tag{1.176}$$

The new vector potential is still a constant field, therefore the forces do not change. Still, the spatial components have a different value than the time ones, therefore the direction of momentum is corrected.

What we find, then, is that Galilean transformations do not preserve the relationship between kinetic momentum and conjugate momentum. To recover the original relationship, one has to perform a gauge transformation. But this is contrary to the expectation that the laws are the same for all inertial frames. There are likely other deeper issues at play, which we hope to uncover in the future.

Metric tensor revisited

As we saw, the metric tensor appears not as defining the distances in space-time but rather as the linear relationship between velocity and conjugate momentum. It does, however, end up playing an important geometric role: in 1.159 we see that the metric tensor times the mass is the off diagonal component of the form ω_{ab}. Therefore, if we have a range of positions dx^α and a range of velocities dv^β, the number of configurations they identify is given by $dx^\alpha m g_{\alpha\beta} dv^\beta$. In other words,

Insight 1.177. *the metric tensor allows us to count states in terms of the kinematic variables.*

This corroborates what we saw in 1.127: the determinant of the metric tensor is the Jacobian determinant that allows us to express densities over phase space as densities over kinematic variables. In relativity, the square root of the determinant of the metric tensor appears to create an invariant volume element. That is, $\int_U \sqrt{|g_{\alpha\beta}|}\, dx^0 dx^1 dx^2 dx^3$ defines the volume of space-time region U. However, we can also define an invariant volume element over the space of kinematic variables. That is, $\int_V |g_{\alpha\beta}|\, dx^0 \cdots dx^3 du^0 \cdots du^3$ defines the volume of the position-velocity space. The determinant of the metric tensor, then, is more directly giving us the size of a volume taken in position-velocity space instead of space-time. This also connects the ability to specify distributions in terms of kinematic variables since, as we saw in FKE-DEN, the Jacobian determinant of the transformation between position and velocity is proportional to $|g_{\alpha\beta}|$.

Why is it, then, that defining volumes over kinematic quantities fixes the geometry of space-time? Note that both velocities and differentials of space transform like vectors since $du^\alpha = dx^\alpha/d\tau$. Therefore if $dx^\alpha g_{\alpha\beta} du^\beta$ is invariant, then $dx^\alpha g_{\alpha\beta} dx^\beta$ will also be invariant. The geometry of space-time, then, is set by the geometry of phase space.

There are other interesting open questions to explore to better understand this relationship. For example, in relativity, one does put spatial and time variables on the same plane. However, we saw that the primary role of the metric tensor is so that we deal with distributions and integration, and these work differently in space and time. Consider, in fact, a distribution $\rho(x^i, v^j, t)$. This would represent the evolution over time of a distribution over kinematic variables. At each time, the distribution has to integrate to one, assuming it is normalized, which means the units of ρ are $\left[\frac{1}{[x]^3[v]^3}\right]$, the inverse of the cube of distance times velocity. Note that every observer has to see the same thing: a normalized distribution at each moment in time. Additionally, this has to happen no matter what ρ is. An open question in reverse physics, then, is whether imposing this requirement is already enough to recover parts of relativity.

1.10 Reversing phase space

In this section we will find the assumptions required to rederive the structure of phase space. The principle of relativity will play a key role as the structure of classical phase space is the only structure that allows us to define densities and entropy in a way that is coordinate invariant.

Properties of phase space

We have seen that DR and IND are the constitutive assumptions of Hamiltonian mechanics, in the sense that they fully characterize Hamiltonian evolution. The addition of KE in its

full version recovers both Lagrangian mechanics and massive particles under potential forces. The relativistic version of the theories comes out without additional assumptions simply by properly dividing the role of time as a variable from time as a parameter. But in all this discussion we assumed, without questioning, that states are identified by pairs of variables: position-velocity or position-momentum. Is this a coincidence or is there an underlying reason for it? Can we find assumptions from which the structure of phase space itself can be recovered?

Since we are interested in the structure of phase space itself, let's go back to its original version, without the extension to time or the temporal DOF. Throughout the whole discussion, conjugate momentum p_i has been written with the index down, tacitly assuming it to be a covector. This means it obeys the following transformation rules under coordinate changes:

$$\begin{aligned} \hat{q}^i &= \hat{q}^i(q^j) \\ \hat{p}_i &= \partial_{\hat{q}^i} q^j \, p_j \end{aligned} \tag{1.178}$$

When a metric tensor is defined, that is when assumption KE is valid, the difference between vector and covector blurs, as we can always transform one into the other, but this is not the general case. We have, therefore, this condition

> Conjugate momentum p_i changes like a covector under changes of coordinates q^i (PS-COV)

which characterizes the relationship between q^i and p_i. We want to stress that these changes of coordinates do not mix space and time variables. Therefore we are only taking different choices of spatial coordinates at equal time.

Given that the form ω_{ab} plays a fundamental role, as it defines the geometry of phase space in terms of state count, let's see how it transforms during a coordinate change. We have:

$$\begin{aligned} \hat{\omega}_{ab} &= \begin{bmatrix} \omega_{\hat{q}^i\hat{q}^j} & \omega_{\hat{q}^i\hat{p}_j} \\ \omega_{\hat{p}_i\hat{q}^j} & \omega_{\hat{p}_i\hat{p}_j} \end{bmatrix} = \begin{bmatrix} \partial_{q^k}\hat{q}^i & \partial_{p_k}\hat{q}^i \\ \partial_{q^k}\hat{p}_i & \partial_{p_k}\hat{p}_i \end{bmatrix} \begin{bmatrix} \omega_{q^kq^l} & \omega_{q^kp_l} \\ \omega_{p_kq^l} & \omega_{p_kp_l} \end{bmatrix} \begin{bmatrix} \partial_{q^l}\hat{q}^j & \partial_{q^l}\hat{p}_j \\ \partial_{p_l}\hat{q}^j & \partial_{p_l}\hat{p}_j \end{bmatrix} \\ &= \begin{bmatrix} \partial_{q^k}\hat{q}^i & 0 \\ \partial_{q^k}\hat{p}_i & \partial_{\hat{q}^i}q^k \end{bmatrix} \begin{bmatrix} 0 & \delta_{kl} \\ -\delta_{kl} & 0 \end{bmatrix} \begin{bmatrix} \partial_{q^l}\hat{q}^j & \partial_{q^l}\hat{p}_j \\ 0 & \partial_{\hat{q}^j}q^l \end{bmatrix} \\ &= \begin{bmatrix} 0 & \partial_{q^l}\hat{q}^i \\ -\partial_{\hat{q}^i}q^l & \partial_{q^l}\hat{p}_i \end{bmatrix} \begin{bmatrix} \partial_{q^l}\hat{q}^j & \partial_{q^l}\hat{p}_j \\ 0 & \partial_{\hat{q}^j}q^l \end{bmatrix} \\ &= \begin{bmatrix} 0 & \partial_{q^l}\hat{q}^i\partial_{\hat{q}^j}q^l \\ -\partial_{\hat{q}^i}q^l\partial_{q^l}\hat{q}^j & -\partial_{\hat{q}^i}q^l\partial_{q^l}\hat{p}_j + \partial_{q^l}\hat{p}_i\partial_{\hat{q}^j}q^l \end{bmatrix} \\ &= \begin{bmatrix} 0 & \partial_{\hat{q}^j}\hat{q}^i \\ -\partial_{\hat{q}^i}\hat{q}^j & -\partial_{\hat{q}^i}\hat{p}_j + \partial_{\hat{q}^j}\hat{p}_i \end{bmatrix} = \begin{bmatrix} 0 & \delta_{ij} \\ -\delta_{ij} & 0 \end{bmatrix} \end{aligned} \tag{1.179}$$

For the last step, we are taking partial derivatives of the new variables with respect to the new variables. The derivative is equal to one if we are taking the partial derivative with respect to the same variable and zero otherwise. We find that condition

> The form ω_{ab} is invariant under changes of coordinates q^i (PS-SYMP)

is implied by PS-COV.

As usual, we ask whether the converse is true. That is, suppose we perform a change of coordinates $\hat{q}^i = \hat{q}^i(q^j)$ for which ω_{ab} is invariant. Does this pose restrictions on how conjugate momentum changes? We have

$$\begin{aligned}
\hat{\omega}_{ab} &= \begin{bmatrix} \omega_{\hat{q}^i\hat{q}^j} & \omega_{\hat{q}^i\hat{p}_j} \\ \omega_{\hat{p}_i\hat{q}^j} & \omega_{\hat{p}_i\hat{p}_j} \end{bmatrix} = \begin{bmatrix} \partial_{q^k}\hat{q}^i & \partial_{p_k}\hat{q}^i \\ \partial_{q^k}\hat{p}_i & \partial_{p_k}\hat{p}_i \end{bmatrix} \begin{bmatrix} \omega_{q^kq^l} & \omega_{q^kp_l} \\ \omega_{p_kq^l} & \omega_{p_kp_l} \end{bmatrix} \begin{bmatrix} \partial_{q^l}\hat{q}^j & \partial_{q^l}\hat{p}_j \\ \partial_{p_l}\hat{q}^j & \partial_{p_l}\hat{p}_j \end{bmatrix} \\
&= \begin{bmatrix} \partial_{q^k}\hat{q}^i & 0 \\ \partial_{q^k}\hat{p}_i & \partial_{p_k}\hat{p}_i \end{bmatrix} \begin{bmatrix} 0 & \delta_{kl} \\ -\delta_{kl} & 0 \end{bmatrix} \begin{bmatrix} \partial_{q^l}\hat{q}^j & \partial_{q^l}\hat{p}_j \\ 0 & \partial_{p_l}\hat{p}_j \end{bmatrix} \\
&= \begin{bmatrix} 0 & \partial_{q^l}\hat{q}^i \\ -\partial_{p_l}\hat{p}_i & \partial_{q^l}\hat{p}_i \end{bmatrix} \begin{bmatrix} \partial_{q^l}\hat{q}^j & \partial_{q^l}\hat{p}_j \\ 0 & \partial_{p_l}\hat{p}_j \end{bmatrix} \\
&= \begin{bmatrix} 0 & \partial_{q^l}\hat{q}^i\partial_{p_l}\hat{p}_j \\ -\partial_{p_l}\hat{p}_i\partial_{q^l}\hat{q}^j & -\partial_{p_l}\hat{p}_i\partial_{q^l}\hat{p}_j + \partial_{q^l}\hat{p}_i\partial_{p_l}\hat{p}_j \end{bmatrix} = \omega_{ab} = \begin{bmatrix} 0 & \delta_{ij} \\ -\delta_{ij} & 0 \end{bmatrix}
\end{aligned} \tag{1.180}$$

The two off diagonal terms impose the same constraint

$$\partial_{q^l}\hat{q}^i\partial_{p_l}\hat{p}_j = \delta_{ij}. \tag{1.181}$$

In matrix terms, we are taking the product of two matrices and equating it to the identity matrix. Therefore the two matrices are the inverse of each other:

$$\partial_{p_l}\hat{p}_j = (\partial_{q^l}\hat{q}^j)^{-1} = \partial_{\hat{q}^j}q^l. \tag{1.182}$$

This means that the change of position variables induces a change of momentum

$$\hat{p}_j = \partial_{\hat{q}^j}q^i\, p_i + A_j(q^k), \tag{1.183}$$

where A_j are arbitrary functions, which we can set to zero without loss of generality.[38] This makes momentum a covector.

The second constraint comes from the bottom diagonal term. Using the newly found transformation rules, we have

$$-\partial_{p_l}\hat{p}_i\partial_{q^l}\hat{p}_j + \partial_{q^l}\hat{p}_i\partial_{p_l}\hat{p}_j = -\partial_{\hat{q}^i}q^l\partial_{q^l}\hat{p}_j + \partial_{q^l}\hat{p}_i\partial_{\hat{q}^j}q^l = -\partial_{\hat{q}^i}\hat{p}_j + \partial_{\hat{q}^j}\hat{p}_i = 0. \tag{1.184}$$

This constraint is satisfied simply because position and momentum are different state variables, and partial derivatives along one variable are taken keeping the others constant. By imposing the preservation of the form ω_{ab} under an arbitrary change of coordinates, we recovered the transformation law of momentum as a covector. Therefore conditions PS-COV and PS-SYMP are equivalent.

The invariance of the form ω_{ab} under coordinate changes means that all those properties that were invariant under Hamiltonian evolution are also invariant under equal-time coordinate changes. Therefore condition

> The Poisson brackets are invariant under equal-time coordinate changes (PS-POI)

[38] The arbitrary functions change the value of zero momentum, potentially at every point in a different way. While this is mathematically possible, it would make no physical sense that a coordinate change would induce a change in the zero reference for momentum. We will see later how this is related to gauge transformations.

is equivalent to PS-SYMP.

The following conditions are equivalent to each other and are implied by PS-SYMP, but do not imply it.

The system allows statistically independent distributions over each DOF under any choice of coordinates q^i	(PSI-DEN)
The system allows informationally independent distributions over each DOF under any choice of coordinates q^i	(PSI-INFO)
The system allows peaked distributions where the uncertainty is the product of the uncertainty on each DOF under any choice of coordinates q^i	(PSI-UNC)

Physically, these correspond to the independence of DOFs, which is assumption IND. The only difference is that the assumption must be valid for all equal-time coordinate changes, which is just an application of the principle of relativity.

Lastly, the following conditions are all equivalent but independent of the PSI conditions and are implied by PS-SYMP, but do not imply it.

Phase space volumes are invariant under equal-time changes of coordinates q^i	(PSV-VOL)
The Jacobian for the transformation induced by equal-time changes of coordinates q^i is unitary	(PSV-JAC)
Densities over phase space are invariant under equal-time changes of coordinates q^i	(PSV-DEN)
Thermodynamic entropy is invariant under equal-time changes of coordinates q^i	(PSV-THER)
Information entropy is invariant under equal-time changes of coordinates q^i	(PSV-INFO)
Uncertainty of peaked distributions is invariant under equal-time changes of coordinates q^i	(PSV-UNC)

Physically, these correspond to requiring that the count of states is the same under all choices of coordinates at a given time. In the same way that conservation of ω_{ab} in time was equivalent to assumptions DR and IND combined, any PSV condition together with any PSI condition will be equivalent to any PS condition. That is, the phase space structure corresponds to invariance of state count plus independence of DOFs.

It should be evident that all these properties are necessary if we want to have a physically meaningful state space. If state count, densities, entropy, independence of DOFs were not properties that all equal-time observers could agree on, there would be no notion of an objective state to begin with, and it would be pointless to even talk about isolation, determinism, thermodynamics and so on. All these properties, then, are constitutive assumptions for any state space, as without them there are no well defined states. Having established that phase space has all these properties, do these properties define phase space? That is, suppose the states of our system are identified by a finite number of continuous quantities, meaning that the state space is a manifold; does the invariance of those properties constrain that space to be phase space?

Invariance over the continuum

To simplify the matter, let us consider the case of a single DOF. One way to define the problem is to ask that if a distribution is uniform for one observer, it should be uniform for all equal-time observers. Uniform distributions are important in statistical mechanics since the macrostate of an isolated system is assumed to be a uniform distribution over all possible microstates that satisfy a few constraints, such as the value of the energy, the number of particles and so on. This is very similar to the "principle of indifference" in classical probability, which assigns equal probabilities to outcomes for which there is no justifiable preference.

For example, if we have a fair die, meaning that the die itself and the mechanism for throwing it do not have a preference for any side, the principle tells us to assign equal probability to all sides.[39] If the die has 6 sides, we assign probability 1/6 for each side. The part of probability theory that is linked to combinatorics works like this. Note that what number we use to label each side of the die, or whether we use something else (e.g. for poker dice, images of playing cards) is irrelevant for the probability assignment. Therefore, when applied to discrete variables, the principle of indifference is invariant under relabeling.

If we try to apply the same idea on the continuum, however, things are not as simple. Suppose we have a factory that produces boxes at random, with side from $1m$ to $3m$. Furthermore, let's assume that the manufacturing procedure does not have a preference for the size of the boxes. Using the principle of indifference, we assume a uniform distribution for the side from $1m$ to $3m$. However, we could have equally said that the factory does not have a preference for the total volume, and assign a uniform distribution for the volume from $1m^3$ to $27m^3$. The problem is that $x \to x^3$ is a non-linear transformation, and uniform distributions do not remain uniform under non-linear transformations. Another way to frame it, on the continuum we do not have a probability, but a probability density, which has units of probability over units of the variable. In the first case the units were probability over meters while in the second probability over meters cubed. The density, then, is unit dependent, it depends on the variable.

This is essentially our problem:

Insight 1.185. *Densities over continuum variables depend on the choice of variable and therefore do not, in general, satisfy the principle of relativity.*

If we have a uniform distribution over a variable, it will only remain uniform under linear changes of variable. This means that density, thermodynamic entropy, information entropy, all the properties we looked at before, are not the same under variable changes. Mathematically, the issue is that only transformations with unitary Jacobian preserve these properties, which is not the general case.

How does phase space solve the problem? Phase space is defined by two variables, q and p. A change of coordinates only changes q arbitrarily. Once the q change is determined, p changes as a covector. Around each point, we have:

$$\begin{aligned} d\hat{q} &= d_q\hat{q}\, dq \\ d\hat{p} &= d_{\hat{q}}q\, dp \end{aligned} \tag{1.186}$$

[39] The principle of indifference is often stated in terms of degree of belief: it is the agent that has no reason to prefer one outcome over the other. We state it in more objective terms: the system and preparation procedure do not have a preference. If an agent believes a die to be fair, while in fact the die is loaded, the principle of indifference gives wrong empirical results, and therefore, for us, it does not apply.

which means

$$d\hat{q}d\hat{p} = d_q\hat{q}\, dq\, d_{\hat{q}}q\, dp = d_q q\, dq\, dp = dq\, dp. \tag{1.187}$$

The area is conserved under a generic change of q precisely because p changes in the opposite way.[40]

While the math is clear, what is the physics behind this? Why does conjugate momentum magically change when we change position? This is better understood if we think in kinematic variables: position and velocity. Why would velocity change when we change position? Because units of velocity are units of position over time. If we redefine units of position, we will redefine units of velocity as well. As we said before, the mathematical tools we currently use do not capture all the physical elements, and unit dependence is something that mathematics currently fails to capture. The defining role of coordinate variables q^i, then, is not that they describe position, but rather the following:

Insight 1.188. *The coordinate variables q^i define the unit system.*

What are the units of conjugate momentum, then? Given that the product $dqdp$ is invariant and it measures the count of configurations for one DOF, p_i must have units of configurations divided by units of the corresponding q^i. By convention, we measure phase-space areas in units of angular momentum, in units of action h, but why do we do that? Suppose that we track position by an angle in radians q^θ. Conjugate momentum p_θ will be expressed in units of area of phase space, and therefore angular momentum. This is the connection between angular momentum and phase-space areas: the conjugate of a dimensionless quantity must have the same physical dimensions as areas of phase space. However, if we change the angle from radians to degrees, conjugate momentum will be units of h over degrees. That is, it is improper to say the count of states is expressed in units of angular momentum. Moreover, while two quantities that represent the same physical quantity must have the same physical dimension, the converse is not true.[41] Ultimately, areas in phase space are measured with those units because, in an inertial Cartesian coordinate frame, linear kinetic momentum times distance can be used to count the configurations of the system, as we saw when discussing the full kinematic equivalence.

To recap, we saw that we cannot define a coordinate invariant density over a single continuous variable, but we can do that over two variables, if one has inverse units with respect

[40] In statistical mechanics, it is often said that Liouville's theorem justifies the use of phase space volume as state count. Liouville's theorem states that under Hamiltonian evolution the density, or equivalently the phase space volume is conserved, which corresponds to assumption DR. As we mentioned before, the conservation of volume in time would not be physically meaningful if the volume were not first an objective quantity. Therefore Liouville's theorem does *not* justify the use of phase space volume as state count: it is the invariance under equal-time coordinate changes that does.

[41] In general, units and physical dimensions keep track of some aspects of physical quantities, but not everything. For example, one can show that, dimensionally, pressure is equal to energy over volume. This relationship is actually useful when describing fluids. On the other hand, energy density is another useful quantity to, for example, measure the capacity of batteries. There is no relationship between the energy density of a battery and its pressure. Another puzzle: if one multiplies radians and meters, what is the resulting unit? Consider the area of the side of a cylinder section $\theta r h$ where θ is the angular size of the section, r is the radius and h is the height of the cylinder. If we multiply θ and r, we get the length of the arc, which is a distance: we get meters. However, if we multiply θ and h, we do not get a length, we have an angle times length: we get meters times radians. The issue is that an angle is really a ratio between the length of the arc and the length of the radius. Only a multiplication by the correct radius will give back a distance.

to the first. Is this the only case? What happens if we use three or more variables? Suppose that q is the only variable that defines the unit system for a set of states, meaning that all other variables have derived units. Then under unit change $\hat{q} = \hat{q}(q)$ the units of all other variables must be uniquely defined. Suppose that we add a single additional constraint, that the measure of states is invariant. How many total variables can the state space have? We saw there must be more than one variable, or we cannot change the unit arbitrarily. Now suppose that we have three or more variables. Given that we have only two constraints, the choice of unit change and the area conservation, these are not enough to fully determine the transformation of all variables, and therefore the units of all variables. Therefore it would not be true that the units of q fully determine the units of all variables. This tells us that, to have densities that are invariant under equal-time coordinate changes, a state space for a system that is fully characterized by a single unit must have exactly one additional quantity, conjugate momentum, that changes covariantly under unit change. We reach the following

Insight 1.189. *Degrees of freedom are two dimensional because only these allow coordinate invariant densities and count of configurations.*

Invariance over multiple degrees of freedom

We saw how things work for a single DOF, let's see how they generalize for multiple independent DOFs. Suppose states are identified by m variables ξ^a, but the unit system is fully identified by a subset q^i of n variables. This would be the case, for example, if we are studying particle trajectories as the spatial variables define the unit system. Moreover, suppose we assume IND, that the system is decomposable into independent DOFs. In this case, independence of the variables ties in with the independence of the DOFs. That is, if the unit of one variable depends on the unit of another, they cannot belong to independent DOFs; on the other hand, if two q^i define independent units, they must belong to independent DOFs by definition.

For example, suppose we have three degrees of freedom. Then we must have three variables q^x, q^y and q^z that define the units of each degree. Suppose we change q^x to $\hat{q}^x$ while leaving the others unchanged. Given the premise and what we discussed above, there must be another variable p_x with inverse units. This must be an additional variable given that q^y and q^z must have independent units. Given that the DOFs are independent, we should also be able to fix the value of both q^x and p_x and obtain a subspace identified by all the remaining variables. Now suppose that we change q^y by itself. By the same logic, there must be another variable p_y. This must have units of inverse q^y, and therefore cannot be any of the variables we already have. We repeat the logic and we find another variable p_z. At this point, there cannot be other variables: we would have to introduce units that do not depend on q^x, q^y and q^z, but this contradicts the assumption that q^i define the unit system. In general, then, if we have n degrees of freedom, we must have $2n$ state variables: the q^i that define the units and the conjugate quantities p_i expressed in inverse units.

We now want to introduce a form ω_{ab} that returns the count of independent configurations for an infinitesimal parallelogram. Here we simply use the same arguments that we gave discussing assumption IND. The number of independent configurations identified by a parallelogram within the same degree of freedom will be given by the area of the parallelogram. On the other hand, a parallelogram across degrees of freedom, for example formed by Δq^x and Δp_y, will not properly identify independent configurations. Note, in fact, that their product

is not unit invariant. This means ω_{ab} must match the familiar form. In short, assumption IND plus requiring that the count of states is the same for all equal-time observers gives us the structure of phase space. That is, condition

> The space allows coordinate invariant distributions over equal-time independent DOFs (PS-INV)

is equivalent to PS-SYMP and therefore to PS-COV and PS-POI.

Phase space and its structure, then, are neither a coincidence nor a choice. Phase space is the only space that allows us to define key concepts, like uniform distributions or reversibility, in an invariant way, which would otherwise be ill-defined over the continuum. In other words:

Insight 1.190. *The structure of phase space is exactly the structure needed to define state densities, thermodynamic entropy, information entropy and statistical uncertainty over continuous quantities in a way that satisfies the principle of relativity for equal-time observers.*

The structure of phase space, then, comes out from simply keeping track of unit dependence between variables. This should sound familiar, as unit dependence was used in condition FKE-UNIT to motivate why assumption KE should be implemented in its full form FKE-LIN. The existence of a metric tensor, of inertial mass, and more could be seen as stemming from assuming that the change of variables from dynamical to kinematic variables depended only on position, on q^i. Now we see that the very structure of phase space rests precisely on the fact that coordinates are those variables that define the units. This leads to the following observation: while taking the weak form WKE-INV of assumption KE leads to no mathematical inconsistencies, it would be physically inconsistent. On one side, when defining phase space, we are saying that the units of the coordinates q^i determine both the units of velocity and of momentum; on the other side, when introducing WKE, we claim that the coordinates q^i do not determine by themselves the unit transformation between velocity and momentum. The fact that we can write a mathematically consistent model that is physically inconsistent is evidence that we need a better mathematical specification of our physical theories. Namely, we need a way to encode unit relationship between variables.

We are also now in a position to answer another question: why are the laws of physics second order? That is, why is it that external forces specify the acceleration of the system, and not the velocity or the jerk (i.e. the derivative of the acceleration)? Note that, once the structure of phase space is derived, we know that, under KE, position and velocity fully determine the state, and, during the evolution, the change of state is fully determined by the acceleration. In other words, the laws of physics are second order exactly because the state is given by variable pairs. Since we have an explanation as to why the state space must obey that structure, that same explanation tells us why the laws are second order. Therefore, we have

Insight 1.191. *The principle of relativity ultimately requires the laws of motion to be second order.*

The principle of relativity, then, is responsible for much more than the invariance of the laws, it is also responsible for the structure of the space and the nature of the laws themselves.

Invariance under relative motion

So far, we have applied the principle of relativity only to transformations that leave time unchanged. For completion, let us apply it to the more usual case, when space and time coordinates are mixed. Consider phase space extended by time only. If we do not mix space and time variables, we keep the equal-time surfaces the same. Therefore we are only requiring that areas, density and entropy over each time slice remain the same. Moreover, we have no requirement to relate what happens on two different surfaces at equal time. That is, we will have the structure of phase space at each time, but there is no connection between these structures at different times. However, if we mix space and time variables, equal-time surfaces for one observer are not necessarily equal-time surfaces for another.

Now suppose we have a distribution that is uniform for one observer. The only way that it is uniform for all other observers is if it remains constant in time as well. In other words, the principle of relativity will imply the invariance of the form ω_{ab} even under space-time transformations, which means, as we saw, Hamiltonian mechanics. The only way that we can satisfy the principle of relativity, then, seems to assume deterministic and reversible motion. That is, condition

> The space allows coordinate invariant distributions over independent and temporal DOFs (DI-INV)

is equivalent to DI-SYMP, and is therefore equivalent to Hamiltonian mechanics.

This may be surprising at first, but in retrospect it makes sense. The invariance among equal-time observers imposed that the density at each state was the same for everybody. In general, however, the equal-time surface for two generic observers will not be the same, but only intersect, in a region. Clearly, on that region the density must be the same for both, but what happens in the regions that are different? Well, if they still need to see the same density distribution, with the same entropy, then we must be able to map one region to the other in a bijective way. That is, the evolution must map each state of one surface to a state of the other surface while preserving the density. But this is exactly what deterministic and reversible evolution does. If we don't have this, if densities spread or if states are not mapped one-to-one, the two observers will see a different state.

As another way to see this, suppose we have a box of gas at equilibrium. Now suppose that at time t_0 we start heating it very slowly, such that we can assume the gas is at equilibrium at each time. The system, then, will increase its entropy. Suppose we stop the process at time t_1, so that the temperature of the gas remains constant after that. Now, the equal-time surface of an observer boosted with respect to the gas will cross the original frame at different times, and therefore will see the gas at different temperatures in different places at the same time. The moving observer will see a system out of equilibrium. So, again, processes that are completely isolated, for which entropy is conserved, are the only processes that are truly relativistic.

This tells us that if we want to study a non-deterministic and/or non-reversible process, in a way that is frame independent, we need to be able to compare it to other processes that can be assumed to be deterministic and reversible. It is only by comparing the two that we are going to be able to ascertain precisely by how much the first one fails to be deterministic or reversible. Suppose, in fact, that we want to construct a coordinate system to study our non-deterministic and non-reversible process. Operationally, this means devising a system of rods and clocks so that we can correlate the quantities measured with our spatial and time references. Therefore we need some guarantee that, as time evolves, the spatial and temporal

references will have set spatial and temporal relationships. For example, if we leave a mark at a particular position, we may expect the mark to remain there; if the clock ticks uniformly in time, we may expect it to do so going forward. But this implicitly requires the existence of at least some deterministic and reversible process that can be used to distribute our references in space and time. If we do not have access to any deterministic process, no reliable coordinate system can be constructed. If the processes are not reversible, references will drift and the precision of our reference system will degrade since DR-UNC is not satisfied.

We close, then, with the following.

Insight 1.192. *Hamiltonian evolution over phase space is exactly the structure needed to define state densities, thermodynamic entropy, information entropy and statistical uncertainty over continuous quantities in a way that satisfies the principle of relativity.*

Gauge transformations

Throughout all this work we have seen that position q^i and momentum p_i are the true state variables: they allow us to identify states uniquely in all circumstances, properly count them and define densities that are invariant under coordinate changes and are defined at a particular instant in time. On the other hand, kinematic variables x^i and v^i identify states only under KE, they properly count states and define densities only in Cartesian inertial frames for systems subject to no forces, and velocity is technically defined over two infinitesimally close instants in time. Yet, under assumption KE, the information content of the two sets of variables are the same. Moreover, position and velocity are the variables that are more directly linked to experimentation: while we can measure velocity and kinetic momentum directly, conjugate momentum is not a direct observable.

We now pose the question: if velocity is the variable we actually measure, is conjugate momentum even uniquely defined? That is, is it possible to change momentum in a way that leaves velocity, and therefore the trajectories, unchanged? Since the position does not change, the new state variables can be written as $(q, \hat{p})$ where the new conjugate momentum $\hat{p}_i = \hat{p}_i(q^j, p_k)$ is a function of the old variables. We want the new variables to be conjugate, therefore the following relationships in terms of the Poisson brackets must hold:

$$\begin{aligned} \{q^i, q^j\} &= 0 \\ \{q^i, \hat{p}_j\} &= \delta^i_j \\ \{\hat{p}_i, \hat{p}_j\} &= 0. \end{aligned} \tag{1.193}$$

Using the second equation we find

$$\begin{aligned} \delta^i_j = \{q^i, \hat{p}_j\} &= \partial_{q^k} q^i \partial_{p_k} \hat{p}_j - \partial_{p_k} q^i \partial_{q^k} \hat{p}_j \\ &= \delta^i_k \partial_{p_k} \hat{p}_j - 0 \partial_{q^k} \hat{p}_j \\ \partial_{p_i} \hat{p}_j &= \delta^i_j. \end{aligned} \tag{1.194}$$

Integrating this last equation yields

$$\hat{p}_j = p_j + G_j(q^i) \tag{1.195}$$

where G_j are arbitrary functions of position. Using the third Poisson bracket, we have:

$$\begin{aligned} 0 = \{\hat{p}_i, \hat{p}_j\} &= \{p_i + G_i, p_j + G_j\} \\ &= \{p_i, p_j\} + \{p_i, G_j\} + \{G_i, p_j\} + \{G_i, G_j\} \\ &= 0 - \{G_j, p_i\} + \{G_i, p_j\} + 0 \\ &= -\partial_{q^i} G_j + \partial_{q^j} G_i \\ &= curl(G_j). \end{aligned} \tag{1.196}$$

This G_j is a curl free field, it admits a scalar potential $f(q^i)$. Therefore the general transformation is of the form

$$\hat{p}_j = p_j + \partial_j f(q^i). \tag{1.197}$$

In other words, conjugate momentum is defined up to a gauge, in the same way that the electromagnetic potential is defined up to a gauge. In fact, these are essentially the same gauges.

Recall the expression

$$u^\alpha = \frac{1}{m} g^{\alpha\beta}(p_\beta - \mathfrak{q} A_\beta). \tag{1.198}$$

If the four-velocity u^α and metric tensor $g^{\alpha\beta}$ are gauge independent, and the electromagnetic four-potential A_β is gauge dependent, then it must be that conjugate momentum p_β must be gauge dependent in the exact opposite way such that their difference is gauge independent. As another way to see this, suppose we have a charged particle in a field free region. In an inertial frame, we would write

$$p_i = m v^i. \tag{1.199}$$

But this not only requires us to use Cartesian coordinates, it also requires us to have chosen the gauge for the magnetic field such that $A_i = 0$. Choosing a gauge for momentum, then, is the same as choosing a gauge for the magnetic field.

If we look again at equation 1.195, note that what the gauge does is redefine the zero momentum at each point. The G_j field represents the new value of the old zero. The new zeros must form a curl free field because the new conjugate momenta still need to form independent DOFs. Recall that when recovering phase space, in equation 1.183, we saw that change of units of position left unspecified a potential change for zero momentum. The gauge freedom is exactly that change.

This tells us, in a more direct way, why assumption KE requires gauge theories. Given that the true observables are the kinematic variables, conjugate momentum is defined only up to a curl-free field that represents the arbitrary definition of zero momentum. Given that the vector potential of the interaction field tells us how the zero momentum is mapped to zero velocity, this inherits the same arbitrariness. The same relationship will be there in quantum mechanics, where the arbitrariness of the zero momentum will be implemented in the arbitrariness of absolute phases.

Entropy and dissipative evolution

Having seen the link between the principle of relativity and deterministic and reversible evolution, let's look a bit more closely at what happens during non-reversible evolution. We saw that a damped harmonic oscillator is deterministic but not reversible, as it does not preserve the count of states. The evolution, in fact, has an equilibrium, and regions around it get smaller and smaller. This presents a problem: if areas get smaller, the entropy goes down. But this is a dissipative system: shouldn't the entropy go up?

To understand the problem, suppose Alice takes a pendulum at rest, displaces the weight and lets it go. Suppose Alice told Bob, "In an hour, the position of the pendulum will be within 1 mm of the equilibrium." This would not give Bob much information as, after an hour, all the energy will have likely dissipated, no matter how it was initialized. Suppose Alice told Bob, "I let the weight go within 1 mm of the equilibrium." This gives Bob a lot more information, as the range of possible trajectories of the pendulum has greatly decreased. Because the motion is irreversible, statements at the same level of precision give more information in the past than in the future.

In terms of information, then, the difference is whether we are asking about how much we know about the position and momentum at a specific time, or how much we know about the system in terms of its overall evolution. If the system satisfies DR, the two are the same. If it doesn't, we have a problem. Intuitively, we would want the state of the system to always provide the same amount of information about the system, but under non-deterministic or non-reversible evolution, this does not work. The amount of information changes in time and, for a boosted observer, it will also change in space and momentum. This brings to light a fundamental problem when defining states and systems: it would seem that such definition can only be given if, at least in some cases, the system can be studied under deterministic and reversible motion.

While the notion of state starts being problematic, the notion of evolution, however, is automatically invariant. A damped harmonic oscillator may not satisfy DR, but it does satisfy KE, which means that each state at a particular time will correspond to a particular trajectory and, given a region of phase space at a particular time, we will be able to quantify the flow of evolutions through that region. As time goes on, the evolutions become more concentrated, and therefore it is more difficult to tell them apart with measurements of the same precision. This is the indication that the system is not reversible. The proper indicator of irreversibility, then, is not the number of states within the region, but the number of evolutions over the region of phase space. This will increase for dissipative systems around the equilibria, and will remain unchanged over reversible evolution.

Note that the flow of evolutions through a surface was exactly the geometric interpretation we found for the principle of stationary action. In the extended phase space, the symplectic form could be understood as quantifying the flow of the displacement field instead of the count of states. Again, this seems to be hinting that the true nature of entropy is not the count of states, but the count of evolutions. We will explore these ideas when applying reverse physics to thermodynamics and statistical mechanics.

1.11 Reversing Newtonian mechanics

In this section we return to Newtonian mechanics, and see that assumption KE is the only assumption that characterizes Newtonian mechanics.

Inertia and forces

We have seen that all systems that satisfy assumptions DR, IND and KE are Lagrangian systems and, therefore, Newtonian given that the acceleration is a function of position and velocity. We also saw that some dissipative systems, like a particle under friction, do not satisfy DR though are Newtonian systems. So now the question is whether IND and KE, by themselves, are enough to characterize Netwonian mechanics.

The first task is to verify that all Newtonian systems satisfy the assumptions. The second law $F^i = ma^i$ in inertial frames makes a link between the dynamics, the force, and the kinematics, the acceleration. Moreover, the force is specified through kinematic variables, which makes it clear that the kinematics is enough to reconstruct the dynamics. Therefore assumption KE is satisfied somewhat trivially because Newtonian systems specify both kinematics and dynamics. Translating the dynamics from forces and masses into energy and conjugate momentum is complicated by the fact that there is no single way to decompose the total force into a conservative and non-conservative part. However, we can always set conjugate momentum equal to kinetic momentum in a Cartesian inertial frame.

As for IND, note that each new variable introduces its own force term, its own velocity and acceleration. The difficulty here is that, since the dynamics does not satisfy DR, this independence may not be kept over time. For example, the forces may be dissipative in such a way that the region with equilibria has lower dimensionality, introducing correlations that are effectively variable constraints. In other words, we do have a notion of independence at each time, but it is not the same notion throughout the evolution. Mathematically, if we are given the map from velocity to conjugate momentum at each time, we would be able to write a form ω_{ab} at each time. Given that the evolution is still differentiable, the Jacobian exists allowing us to map the count of configurations, and the form, back and forth in time. Since the evolution does not, in general, satisfy DR, the form is not conserved over time.

Having seen that Newtonian systems satisfy IND and KE, we have to show the converse, that all systems that satisfy those assumptions are Newtonian systems. We will need all the insights we gained in the previous section. We saw that the principle of relativity mixed with assumption IND recovers the structure of phase space. Assumption KE requires that the kinematics is specified by the dynamics, therefore the state at each time is enough to identify the trajectory of the system. This means that position and velocity must be invertible functions of position and momentum. This also means that the count of configurations can be calculated in terms of regions of position and velocity, which therefore must be well defined over time. Given that finite regions are mapped to finite regions, the evolution must be differentiable, an acceleration must be well defined and must be a function of position and velocity.

Given that position determines the unit system, condition FKE-UNIT is satisfied which is equivalent to FKE-INER. Therefore we find the existence of locally inertial systems, in which a system not subjected to forces will travel in uniform linear motion. This recovers the first law. We already saw that acceleration was fully specified by position and velocity. The mass is the coefficient that recovers the correct count of states, which will also fix the units for the force. This recovers the second law.

Note that, in principle, a more complete account of Newtonian mechanics could be given. This has not been pursued given how Lagrangian and Hamiltonian mechanics play a much bigger role in field theories and in quantum mechanics.

1.12 Directional degree of freedom

When recovering the structure of phase space, we only relied on the premise that q^i define the units. We are now going to use that premise to define a directional degree of freedom, that is a degree of freedom that identifies a direction in space. We will see that, since a DOF must be composed of two variables, the only directional DOFs that are possible are those in three dimensional space.

Magnetic dipole

Spatial and temporal DOFs are of fundamental importance given that every object must be located in space and time. We now want to focus our attention on a different type of DOF, one that captures a direction in space. In general, objects are not completely spherically symmetric and can be distinguished by their orientation. This is also true for fundamental particles, as their intrinsic angular momentum, their spin, gives them an orientation that can be detected and manipulated through magnetic forces.

To study this case, let's consider a magnetic dipole. The magnetic moment μ_i can be written as

$$\mu_i = \frac{\mathfrak{q}}{2m} L_i, \tag{1.200}$$

where $\mathfrak{q}$ is the electric charge, m is the mass and L_i the angular momentum. The Hamiltonian for a magnetic dipole subjected to a magnetic field B^i is given by

$$H = -\mu_i B^i = -\frac{\mathfrak{q}}{2m} L_i B^i. \tag{1.201}$$

To write the equations of motion, we need to know what the conjugate variables of L_i are or, equivalently, the form ω_{ab} or the Poisson brackets. Typically, one is given the Poisson brackets, therefore we will start from those. They are

$$\{L_i, L_j\} = L_k \epsilon_{ijk}. \tag{1.202}$$

In Hamiltonian mechanics, we can write the evolution of any quantity $f(q^i, p_i)$ as

$$d_t f = \partial_{q^i} f d_t q^i + \partial_{p_i} f d_t p_i = \partial_{q^i} f \partial_{p_i} H + \partial_{p_i} f (-\partial_{q^i} H) = \{f, H\}. \tag{1.203}$$

Therefore we have

$$\begin{aligned} d_t L_i &= \{L_i, H\} = \left\{L_i, -\frac{\mathfrak{q}}{2m} L_j B^j\right\} = -\frac{\mathfrak{q}}{2m} B^j \{L_i, L_j\} = -\frac{\mathfrak{q}}{2m} B^j L_k \epsilon_{ijk} \\ d_t \vec{L} &= -\frac{\mathfrak{q}}{2m} \vec{B} \times \vec{L}. \end{aligned} \tag{1.204}$$

Without loss of generality, we can assume that the magnetic field is oriented along the z axis. We have

$$\begin{aligned} d_t L_x &= -\frac{\mathfrak{q}}{2m} B^z L_y \epsilon_{xzy} = \frac{\mathfrak{q}}{2m} B^z L_y \\ d_t L_y &= -\frac{\mathfrak{q}}{2m} B^z L_x \epsilon_{yzx} = -\frac{\mathfrak{q}}{2m} B^z L_x \\ d_t L_z &= -\frac{\mathfrak{q}}{2m} B^z L_i \epsilon_{zzi} = 0. \end{aligned} \tag{1.205}$$

If we integrate, assuming $L_i = L_i^0$ at time $t = 0$, we have

$$\begin{aligned} \omega &= -\frac{\mathfrak{q}}{2m} B^z \\ L_x(t) &= L_x^0 \cos\omega t - L_y^0 \sin\omega t \\ L_y(t) &= L_x^0 \sin\omega t + L_y^0 \cos\omega t \\ L_z(t) &= L_z^0 \end{aligned} \tag{1.206}$$

which gives us the Larmor precession of the magnetic moment. This is clearly a Hamiltonian system, it obeys Hamilton's equations, but it seems to have three variables, the components of L_i, instead of an even number. Shouldn't we have conjugate pairs?

Note that the norm L of the vector is a constant of motion, only the angle changes. We can then write $L_i = Ln_i$ where n_i is a unit vector. The state space can be understood as the space of all possible vectors with the same magnitude, which is the surface of a 2-sphere. The size of an area A over the surface of a sphere of radius r can be measured using the solid angle

$$\Omega = \frac{A}{r^2}, \tag{1.207}$$

which is already a unit independent quantity. The infinitesimal solid angle can be written in terms of the polar angle φ and the azimuthal angle θ

$$d\Omega = \sin\varphi d\varphi d\theta. \tag{1.208}$$

The two variables are not conjugate, as the area is not simply the product of the differentials. However, we can write

$$d\Omega = d(-\cos\varphi)d\theta. \tag{1.209}$$

The $\cos\varphi$ is simply the component along the z direction, while θ is the angle on the x-y plane. Changing the orientation of the solid angle, and relabeling for clarity $\theta = \theta^{xy}$ we have

$$d\Omega = dn_z d\theta^{xy}, \tag{1.210}$$

which means that the angle on a plane and the component orthogonal to the plane are conjugate. If we pick the variable θ^{xy} as the coordinate q and L_z as the conjugate p, we have two variables with the right geometric relationship and with the right units. Therefore

$$\begin{aligned} \omega_{\theta^{xy} L_z} &= -\omega_{L_z \theta^{xy}} = 1 \\ \{\theta^{xy}, L_z\} &= -\{L_z, \theta^{xy}\} = 1 \end{aligned} \tag{1.211}$$

We can recover L_i with the following expressions

$$\begin{aligned} L_x &= \cos\theta^{xy}\sqrt{L^2 - L_z^2} \\ L_y &= \sin\theta^{xy}\sqrt{L^2 - L_z^2} \\ L_z &= L_z. \end{aligned} \tag{1.212}$$

We can also recover the Poisson brackets for the components

$$\begin{aligned}\{L_x, L_y\} &= \partial_{\theta^{xy}} L_x \partial_{L_z} L_y - \partial_{L_z} L_x \partial_{\theta^{xy}} L_y \\ &= -\sin\theta^{xy}\sqrt{L^2 - L_z^2}\sin\theta^{xy}\left(\frac{-L_z}{\sqrt{L^2 - L_z^2}}\right) \\ &\quad - \cos\theta^{xy}\left(\frac{-L_z}{\sqrt{L^2 - L_z^2}}\right)\cos\theta^{xy}\sqrt{L^2 - L_z^2} \\ &= \sin^2\theta^{xy} L_z + \cos^2\theta^{xy} L_z = L_z\end{aligned} \tag{1.213}$$

$$\begin{aligned}\{L_y, L_z\} &= \partial_{\theta^{xy}} L_y \partial_{L_z} L_z - \partial_{L_z} L_y \partial_{\theta^{xy}} L_z \\ &= \cos\theta^{xy}\sqrt{L^2 - L_z^2} - 0 = L_x\end{aligned} \tag{1.214}$$

$$\begin{aligned}\{L_z, L_x\} &= \partial_{\theta^{xy}} L_z \partial_{L_z} L_x - \partial_{L_z} L_z \partial_{\theta^{xy}} L_x \\ &= 0 - (-\sin\theta^{xy}\sqrt{L^2 - L_z^2}) = L_y\end{aligned} \tag{1.215}$$

We can verify that Hamilton's equations work as before. If the magnetic field is aligned along the z direction, we have

$$\begin{aligned}H &= -\frac{\mathfrak{q}}{2m} L_z B^z \\ d_t\theta^{xy} &= \partial_{L_z} H = -\frac{\mathfrak{q}}{2m} B^z \\ d_t L_z &= -\partial_{\theta^{xy}} H = 0.\end{aligned} \tag{1.216}$$

Given initial conditions θ_0^{xy} and L_z^0, we have

$$\begin{aligned}L_z(t) &= L_z^0 \\ \theta^{xy}(t) &= \omega t + \theta_0^{xy} \\ L_x^0 &= \cos(\theta_0^{xy})\sqrt{L^2 + (L_z^0)^2} \\ L_y^0 &= \sin(\theta_0^{xy})\sqrt{L^2 + (L_z^0)^2} \\ L_x(t) &= \cos\theta^{xy}(t)\sqrt{L^2 - L_z^2(t)} = \cos(\omega t + \theta_0^{xy})\sqrt{L^2 - (L_z^0)^2} \\ &= \cos(\omega t)\cos(\theta_0^{xy})\sqrt{L^2 - (L_z^0)^2} - \sin(\omega t)\sin(\theta_0^{xy})\sqrt{L^2 - (L_z^0)^2} \\ &= L_x^0\cos\omega t - L_y^0\sin\omega t \\ L_y(t) &= \sin\theta^{xy}(t)\sqrt{L^2 - L_z^2(t)} = \sin(\omega t + \theta_0^{xy})\sqrt{L^2 - (L_z^0)^2} \\ &= \sin(\omega t)\cos(\theta_0^{xy})\sqrt{L^2 - (L_z^0)^2} + \cos(\omega t)\sin(\theta_0^{xy})\sqrt{L^2 - (L_z^0)^2} \\ &= L_x^0\sin\omega t + L_y^0\cos\omega t\end{aligned} \tag{1.217}$$

which recovers the precession.

Let us sum up what we learned. The magnetic dipole is described by a single DOF, where the conjugate variables are the angle on a plane and the component of the angular momentum perpendicular to the plane. A constant force does not correspond to a constant angular acceleration, but to a constant angular velocity. Now the question is how and to what extent this can all be generalized.

Generalizing to directional quantities

The above treatment works without modification regardless of whether the magnetic dipole describes a small solenoid, a small rotating charge distribution, or the spin of a single particle. Spin was originally thought to be due to a rotational motion of a particle, hence the name. However, this would require, for example, an electron to be spinning at velocities that would exceed the speed of light, therefore this view is no longer favored. But if spin is not due to a rotation, why does it have the same properties as angular momentum? Is it a coincidence?

Let us call directional quantity a quantity l_i defined only by a direction in space, meaning that the magnitude of the quantity $|l_i| = l$ is fixed, an intrinsic feature of the object, while the direction can change. A magnetic dipole, generated by either spin or rotating charge, can be considered a directional quantity. The question now is whether the directional nature is enough to recover its properties.

If we want to keep track of a direction in space, the angle θ^{xy} of the component of the direction on a plane is a natural variable to use. Therefore θ^{xy} is a natural choice for a unit variable. Because of the structure of phase space, we will have a conjugate quantity L_z, that we will call directional momentum. Given that a directional quantity is fully determined by a direction in space, the count of possible configurations is proportional to a solid angle. As we saw before in the context of a magnetic dipole, this means that the conjugate quantity L_z must be proportional to the component of the directional quantity perpendicular to the plane where θ^{xy} is defined. Therefore $l_z = kL_z$ for some constant k.

Recall now that the product of conjugate quantities must always be in units of phase space. We already noted, in fact, that if a coordinate is dimensionless, then its conjugate must have dimensions of angular momentum. Therefore directional momentum L_z must have dimension of an angular momentum, regardless of what the directional quantity l_i is about. Thus, even if the directional quantity is not "an amount of rotation" defined by an angular momentum, we can still write $l_i = kL_i$, where L_i is expressed in units of angular momentum. That is, a directional momentum L_i arises naturally every time we have a directional quantity. In the case of a magnetic dipole $k = \frac{\mu_i}{L_i} = \frac{\mathfrak{q}}{2m}$ is the gyromagnetic ratio.

We want to stress that for a directional quantity both variables, the directional momentum L_z along a given direction z and its angle θ^{xy} on the perpendicular plane, are effectively tracking a single quantity. This is different from what happens in a spatial DOF, where the coordinate q determines position while the conjugate p determines conjugate momentum, and therefore the velocity, which are different properties of the system. This fact tells us that a force acts differently in the two cases. A force for a spatial DOF will, in general, affect momentum, and therefore velocity, which means it will impart an acceleration on position. However, for a directional quantity the only thing that can be changed is a direction, therefore a force can only impart a change in direction, a directional velocity. To sum up, a force over position will impart a constant spatial acceleration while a force over a direction will impart a constant directional velocity.

For a rotating rigid body, the situation may seem different. We do have an angle and an angular velocity, and the angular momentum is the angular velocity times the moment of inertia, much like kinetic momentum is mass times velocity. Given that angular momentum is conserved, angular velocity is conserved and a force needs to be used to change the angular velocity. Therefore, at first glance, the situation seems different from the magnetic dipole. On the other hand, we are dealing with an angular momentum in both cases, therefore the

situation can't be that different.

The issue here is that there are two directions: one is the direction of the axis of rotation, the other is the orientation of the object on the plane of rotation. While the second one constantly changes, the first may be fixed. Note that the ability to discern the second assumes the ability to discern parts of the rigid body. If the object is too small with respect to our resolution, the orientation angle cannot be defined. On the other hand, even if we study the overall object, we would still be able to discern the direction and magnitude of the angular momentum by performing a rotation. In fact, if the rotation is performed along a direction different from the direction of the angular momentum, a torque will need to be applied to perform the rotation. The greater the angular momentum, the greater the torque. We are now in a case similar to the magnetic dipole, where a constant force is needed to impart a constant velocity.

We have seen that if we want to describe a directional quantity, and only a directional quantity, with state variables, there is only one way to do it, and it will lead to a notion of directional momentum. Spin and angular momentum are, in our definition, two types of directional momentum. Is this the only way to do it? That is, why does a directional DOF only includes the directional quantity, without the directional velocity?

Suppose we want to describe a particle, in the sense of an infinitesimally small part of an object, which is not only characterized by a position, but also by a directional quantity. If we pick a specific time t, the position and direction at that particular time are independent variables, in all senses: we can choose units independently, we can choose from all possible configurations of each variable and a choice for one variable does not constrain the choice for another variable. It is true that we will need to relate the direction identified by the directional variable to the reference frame used by the position coordinates. Yet, we are not forced to choose a particular plane for the directional angle θ.

The situation between velocity and the directional velocity, however, is different. Contrast the following two cases. In the first, we are in an inertial frame with a particle at rest, whose directional quantity changes with a constant directional velocity. In the second, we are in a rotating frame with a particle at the center, whose directional quantity changes with a constant directional velocity. If we look only at the directional quantity, these two cases are indistinguishable: we need to know which frame we are in. But knowing the frame, ultimately, means knowing the definition of our spatial variables. That is, there is a definitional dependency between the spatial variables and the directional variables, which are, in the end, defined in space. We are free to choose the relationship between the variables at a specific time. But to relate the directional quantity at different times, we must know how the spatial coordinates change in time. Therefore, the directional quantity is an independent quantity with respect to the spatial DOFs, but the directional velocity is not. We reach the following:

Insight 1.218. *An independent directional DOF must be restricted to a directional quantity.*

We concluded that an independent directional DOF only includes a directional quantity, so the structure we have for the magnetic dipole is an instance of a general structure. However, note how the directional quantity fits exactly in one DOF. Is this necessary? That is, could we have a directional quantity that fits in more than one pair of conjugate quantities? In other words, is it possible to have directional quantities that break up into multiple independent DOFs?

As we said before, identifying a direction in three dimensions is equivalent to identifying a point on a 2-sphere. If we had an n-dimensional space, a direction would correspond to a point on an $(n-1)$-sphere. That sphere would have to allow a description in terms of conjugate variables with a suitable ω_{ab}. It is a result of symplectic geometry that the 2-sphere is the only sphere to allow that structure. Therefore

Insight 1.219. *An independent directional DOF can only exist in a three dimensional space.*

In other words, three dimensional space is special as it is the only space that allows us to talk about directions as forming an independent degree of freedom. This is yet another example of how simple unassuming premises have significant consequences.

1.13 Infinitesimal reducibility

We have seen that the structure of phase space (i.e. the symplectic structure) is exactly the structure needed to describe density, count of states and entropy in a coordinate invariant way. This implicitly assumes that states are points on a differentiable manifold, that is they are fully identified by real variables that only allow differentiable changes of variables. We will see that this corresponds to the assumption of infinitesimal reducibility.

Particles as infinitesimal parts

At this point, we have ample evidence that the correct fundamental object that describes classical systems is not a point on phase space. All physical objects have finite spatial extent and assuming otherwise leads to problems. A finite mass concentrated at a point is incompatible with general relativity, as it would puncture space-time and not travel along geodesics,[42] and with quantum mechanics, as it would imply uniform spread in momentum thus requiring infinite kinetic energy.[43] The correct physical object, instead, is a distribution over phase space. It explains the very structure of phase space (i.e. the distribution has to be frame invariant), it explains why the laws of motion are differentiable (i.e. they transport a density, not just points) and why they follow Hamiltonian mechanics for isolated, i.e. deterministic and reversible, systems. We can therefore conclude that

> The state of a classical system is given by a distribution over phase space. (IR-DIST)

As usual in reverse physics, we want to find equivalent physical assumptions for the same statement.

While point particles should not be considered fundamental classical objects, the whole of classical mechanics assumes them and uses them. Therefore we should understand exactly what they represent.

One way to understand point particles is as an approximation. When we calculate the trajectory of the earth around the sun, the motion of a cannonball or the average displacement of a molecule under Brownian motion, we know that the object is not actually point-like. But

[42] Mathematically, it would mean that all mass-energy would be concentrated in points of infinite curvature, leaving the rest of space-time flat with no mass.

[43] Mathematically, the wave-function would be a δ-function, which is not even an element of the Hilbert space L^2.

since the body is rigid, and assuming that the effect of external forces on the different parts of the body can be neglected, we can get away with studying the motion of the center of mass. Clearly, we need to check that the assumption holds during the whole motion: if the distance between a comet and the sun becomes smaller then the radius of the sun, the sun cannot be assumed to be point-like anymore. The approximation will also fail if the force exerted on different parts of a body is greater than the force that keeps it together.

The point particle approximation is not unique to classical mechanics. Quantum mechanics, through the Ehrenfest theorem, also allows for a point particle approximation, which will fail if the wave-function is spread over a region where the potential changes are non-negligible. Since the approximation is not unique to classical physics, this characterization of a point particle does not help us understand the fundamental constitutive assumptions of classical mechanics.

When discussing the nature of Hamiltonian evolution, however, we saw that states were better understood as infinitesimal regions of phase space. This means that classical particles, at least this more fundamental version of them, can't be understood as a standalone physical object: they should really be thought of as the limit of some recursive subdivision. That is, classical particles should really be thought of as an infinitesimal part of something bigger, and the 'part' in 'particle' should be understood literally.

The actual classical object, then, is not the infinitesimal part, but the whole object that can be, in principle, divided into infinitesimal parts. In the same way that, when discussing differential topology, we had quantities over finite regions that could be broken into the infinitesimal contributions, the whole object, together with its notion of state, can be understood as made up of infinitesimal parts.

> A classical system can be thought of as being made of infinitesimal parts, called particles. (IR-INF)

All classical systems follow the assumption, not just mechanical systems. Continuum mechanics and fluid dynamics assume that the materials are a continuum of infinitesimally small parts; classical electromagnetism assumes that the electromagnetic radiation can be decomposed into arbitrarily small signals at all frequencies. Moreover, the assumption clearly fails in quantum mechanics. In that case, we cannot talk about the state of a part of an electron; materials are not a continuum of infinitesimally small parts; the intensity of electromagnetic radiation cannot be made arbitrarily small.

Divisible vs reducible

Before going forward, we need to be clear as to what "being made of" actually means. For example, if we say that a table is made of a horizontal top and four legs, we may mean that we can take the table apart and study its components independently, or that we can describe the whole table by describing the top and the legs. While often one can do both, these are actually different properties. That is, divisibility, the ability to "cut" an object into independent parts, is not the same as reducibility, the ability to describe an object in terms of parts. Let us go through some examples.

Suppose we have a planarian worm and we divide it in half. After some time, the tail will regrow a head, and the head will regrow a tail. We will be left with two worms. A planarian worm is divisible into two worms, in the sense that we have a process by which a worm is

divided into two worms, but it is not reducible to two worms, in the sense that describing one worm is not the same as describing two worms.

Now, suppose we have a magnet. We can describe it by describing its north and south pole. Now suppose we divide it in half. Each half will be a new magnet with its own north and south pole. A magnet is reducible to a north and a south pole, meaning that describing a magnet is the same as describing its poles, but it is not divisible into a north and a south pole, meaning that we do not have a process that can separate the two.

We can also find differences between divisibility and reducibility with fundamental particles. Suppose we have a muon. If we wait some time, it will decay into an electron, a neutrino and an antineutrino. A muon divides itself into the three particles, but it is not reducible to the three particles: the state of a muon is not equivalent to the state of an electron and two neutrinos.

Suppose we have a proton. This is not a fundamental particle and it is described in terms of quarks and gluons. However, if we try to separate one of its quarks, the interaction energy will create new quark-antiquark pairs and the proton will divide into multiple hadrons. A proton is reducible to quarks and gluons, but it is not divisible into them.

Divisibility here means that we have a physical process that starts with the overall system and ends with parts that can now be independently manipulated. Mathematically, if S_C is the state space of the overall system and S_A and S_B are the state spaces of the parts, we have a time evolution map $U_t : S_C \to S_A \times S_B$ that takes an initial state of the composite and returns a state for each part. This is not what we want.

Reducibility means that describing the whole is the same as describing the parts. Mathematically, the state space of the whole system $S_C = S_A \times S_B$ is exactly the Cartesian product of the parts. In other words, what we are taking apart is not really the object itself, but its state. This is the type of partitioning we are going to use.

To make reducibility more concrete and intuitive, suppose we have a ball. We can throw the ball, study the motion of the whole ball. We can also take a red marker, make a red dot on the ball, and study the motion of the red dot. We say the ball is reducible because studying the motion of the whole ball is equivalent to studying the motion of all possible red dots. It is infinitesimally reducible if we assume that we can make the red dot arbitrarily small. This is the property we are interested in. Conversely, suppose we have an electron. We can study the motion of the electron, but we cannot make a red dot on the electron. We have no process at our disposal that can tag part of an electron so that we interact with and study only that part. Whenever we interact, we interact with the whole electron. This means that the assumption does not hold in that case.

Condition IR-INF, then, is a property of classical systems and only classical systems, and it is clear that IR-DIST implies IR-INF. But is the converse true? That is, if something can be thought of as being made of infinitesimal parts, does it follow that the state of those infinitesimal parts should be represented by points in phase space? In other words, is condition IR-INF enough to characterize classical systems?

Classical systems and infinitesimal reducibility

Suppose we have a system that satisfies condition IR-INF. Under this premise, we have two state spaces: S_C for the full system and S_P for the infinitesimal parts, the particles. Each state for the full system must tell us exactly how much of the system is in each region of S_P. That

is, for each state $s \in S_C$ we have an associated real valued set function[44] $f(U) \in [0,1]$ that for every region $U \in S_P$ returns the fraction of the system in that region of particle state space.[45] Moreover, reducibility implies f to be additive, meaning that if U is the disjoint union of two regions U_1 and U_2, then $f(U) = f(U_1) + f(U_2)$. Because we are assuming infinitesimal reducibility, this must hold not just in the finite case, but also in the countable case. Mathematically, f is a bounded measure. Since the state of the whole system is fully identified by the state of the parts, for each state $s \in S_C$ there is one and only one f. We can understand S_C as the space of such functions.

Depending on whether we are considering a specific instance of a particular system or an ensemble of similarly prepared systems, the value of the function f will have a different physical meaning. It could be understood as the fraction of a single system that has a particular property (i.e. half of the ball is to the right of the line), or it could be understood as the probability that a particular instance of an ensemble has a particular property (i.e. half the time the ball is to the right of the line). From now on, we will assume we are talking about an actual system, but all that we say will apply to the ensemble case as well.

The fact that $f(U)$ is real valued is also implied by condition IR-INF since the parts are infinitesimal and therefore there can't be a smallest increment.[46] It also implies that the fraction $f(\{x\})$ associated to a single particle state $x \in S_P$ must be zero. Otherwise we would be associating a finite non-zero fraction to an infinitesimal part x, which would make it not infinitesimal. The particle state space S_P of an infinitesimally reducible system, then, must be charted by continuous variables.[47]

So far, we have found that, under condition IR-INF, the state space S_C of a system is the space of all possible functions $f(U) \in [0,1]$ where $U \subseteq S_P$ is a subset of all the possible particle states. Furthermore, the particle state space S_P must be charted by continuous variables, it must be a manifold. We now need to recover the notion of differentiability and of invariant density.

To be able to fully characterize the state of the whole system by the state of the parts, we need to be able to quantify the fraction of the system for each particle state. We saw that the fraction for each particle state is zero, which makes sense because particle states are limits. Therefore the fraction associated to a particle state should be a limit as well: it should be a fraction density, similar to mass or charge density at a point. A fraction density will quantify the fraction present over a region of particle states, meaning that is expressed as a fraction over count of particle states. We therefore need to be able to quantify how many states there are in each region of S_P; we must be able to say which regions have equal number of states. Over a discrete space this would be trivial, we would just count the number of points, but over the continuum finite ranges have infinitely many points and counting points

[44] A set function is a function that takes sets as an argument.

[45] Technically, the region U must be a Borel set, as these are the regions that are associated with an experimental procedure. These details are established in the physical mathematics part of the book.

[46] Technically, this only limits the output of $f(U)$ to be a bounded dense linear order. However, the function f must be closed under arbitrary countable addition, bringing in all the limits, meaning that the ordering must be complete. Lastly, the whole system can always be divided into countably many pieces, which tells us that the order must have a countable dense subset. Since the order is bounded, dense, complete and has a countable dense subset, it is order isomorphic to $[0,1] \subset \mathbb{R}$.

[47] Again, technically we only found that the space must be, in some sense, dense. To show that it is "dense like the reals", we need the link between topology and experimental verifiability that we establish in physical mathematics.

doesn't work. We must have another set function $\mu(U)$, which will also be countably additive as disjoint regions will identify entirely different states, making μ another measure. However, μ will be bounded only from below (i.e. no set can have fewer than zero states) as S_P can have potentially infinitely many possible states.

Note that, because of the infinitesimal reducibility assumption, we will be able to find smaller and smaller subsets of S_P. The value of μ will keep decreasing and, in the limit of infinitesimal subdivision, we will have $\mu(\{x\}) = 0$. That is, a single particle state counts as zero in terms of number of states. In a way, the math is already telling us that particle states do not exist literally and that the assumption of infinitesimal reducibility shouldn't be taken literally. It is a simplifying assumption and should be understood as such. In the same vein, as we can define distributions over phase space that have arbitrarily narrow spread, the entropy of those distributions can be an arbitrarily low number, which will tend to $-\infty$ in the limit of a δ-distribution. This is clearly a problem, since thermodynamics requires the entropy to be non-negative. These two problems are in fact the same problem, since the entropy for a uniform distribution is the logarithm of the count of states. In other words, we shouldn't be surprised that classical mechanics fails for small objects (i.e. narrow distributions), or for ensembles with low entropy: it is exactly when condition IR-INF fails.

Continuing our derivation, we have seen that both the measure f for the fraction of the system and the measure μ for the count of particle states are zero for sets with single points. More in general, we must have that whenever $\mu(U) = 0$, then $f(U) = 0$. That is, we cannot assign a finite fraction to a set of states that has no finite state count. That would, again, mean that we have a finite fraction associated to an infinitesimal subdivision, which is physically untenable. Mathematically, this means that f is absolutely continuous with respect to μ, and therefore we can define a density $\rho = d_\mu f$ such that $f(U) = \int_U \rho d\mu$.[48] That is, the state of the whole system can be described by a density over the particle states.

As we said, S_P is a manifold. The measure μ is a feature of that manifold, in the sense that it depends only on the properties of the particle states, and it will be the same for all states of the whole system. Of all possible state variables that we can use to chart the manifold, then, it is convenient to choose those that can express μ as a density of states over units of the state variables. Mathematically, we are asking that the Lebesgue measure induced by the variable $\xi : S_P \to \mathbb{R}$ is absolutely continuous with respect to μ. This will also require that the variables are differentiable with respect to each other: it will lead to the differentiable structure. That is

Insight 1.220. *The differentiable structure of a state space is exactly the ability to express state count and fractions as densities over said states.*

Lastly, both f and μ must not depend on the choice of state variables. Therefore the proper expression of $\rho = d_\mu f$ must also not depend on the choice of state variables. This, plus assumption IND, as we have seen, requires the existence of the form ω, and recovers phase space, the symplectic manifold. Therefore we have recovered IR-DIST.

Condition IR-DIST is therefore equivalent to condition IR-INF. Classical mechanics describes exactly those systems that follow

Assumption IR (Infinitesimal Reducibility)**.** *The state of the system is reducible to the state of its infinitesimal parts. That is, specifying the state of the whole system is equivalent*

[48] Mathematically, ρ is the Radon-Nikodym derivative.

to specifying the state of its parts, which in turn is equivalent to specifying the state of its subparts and so on.

Assumption IR is, therefore, the constitutive assumption of classical mechanics. Assumptions IR, IND and DR are the constitutive assumptions of the Hamiltonian formulation. Assumptions IR, IND and KE are the constitutive assumptions of Newtonian mechanics. Assumptions IR, IND, DR and KE are the constitutive assumptions of Lagrangian mechanics.

We have found all the constitutive assumptions of all formulations of classical mechanics. We now are guaranteed that every result of classical mechanics is, one way or the other, explained by those assumptions. These are the only physical ideas that are strictly required to understand all the general aspects of classical mechanics. There is nothing else. Therefore the goal of reverse physics for classical mechanics is reached.

1.14 Classical uncertainty principle

Before concluding, let us turn again to the problem of zero measure on individual particle states. As we said, a distribution over phase space can be made arbitrarily narrow and, in the limit of a δ-function, the support will be a single point. Since the entropy is the logarithm of the phase-space volume, which is zero for a single state, the entropy of a δ-function is minus infinity.

Negative entropy is not compatible with the third law of thermodynamics, which states that the entropy must be non-negative. Therefore, on thermodynamics grounds, the spread of a distribution cannot be made infinitesimal. This means that any distribution that is physically meaningful must have a finite spread in position and momentum. Conceptually, this sounds very close to the uncertainty principle of quantum mechanics, so it's worth looking at it more closely.

First of all, since we saw that units have an important role, consider the expression for thermodynamic entropy $S = k_B \log W$ where k_B is the Boltzmann constant and W is a volume of phase space. Given that a logarithm must take pure numbers, the expression is dimensionally incorrect. To correct it, we should take

$$S = k_B \log \frac{W}{h} \tag{1.221}$$

where h can be understood as the volume for which the thermodynamic entropy is zero.[49] As we saw, units of momentum are units of configurations, of areas in phase space, over units of position. Therefore we can choose units of momentum such that h is, numerically, one, but not dimensionally. To keep track of the value and units, let $[qp]$ denote the unit of the product between position and momentum, and let (W) of (h) denote the respective numeric values. We have

$$S = k_B \log \frac{W}{h} = k_B \log \frac{(W)[qp]}{(h)[qp]} = k_B \log \frac{(W)}{(h)} = k_B \log(W) - k_B \log(h) \tag{1.222}$$

where $\log(W)$ is the logarithm of the numeric value of W. This clearly show that changing (h), the numeric value of h, changes the zero for the entropy and therefore the value of h can be fixed by requiring that the entropy is zero if $W = h$.

[49] We stress that here h is defined purely from the zero of the thermodynamic entropy, and it would be a value that is to be found experimentally. Experimentally, one would find the value of the Planck constant.

To make the Gibbs-Shannon entropy consistent with the previous definition, we must set

$$I[\rho] = -k_B \int \rho \log(h\rho)\, dqdp. \tag{1.223}$$

If we take a uniform distribution ρ_U over a region U with volume W, we find

$$\begin{aligned} I[\rho] &= -k_B \int \rho_U \log(h\rho_U)\, dqdp = -k_B \int_U \frac{1}{W} \log \frac{h}{W}\, dqdp \\ &= k_B \log \frac{W}{h} \frac{\int_U dqdp}{W} = k_B \log \frac{W}{h} \end{aligned} \tag{1.224}$$

Similarly to before, we find

$$\begin{aligned} I[\rho] &= -k_B \int \rho \log(h\rho)\, dqdp = -k_B \int \rho \log\left((h)[qp](\rho)[q^{-1}p^{-1}]\right)\, dqdp \\ &= -k_B \int \rho \log(\rho)\, dqdp - k_B \log(h) \\ &= -k_B \int \rho \log(\rho)\, dqdp - k_B \log(h) \end{aligned} \tag{1.225}$$

which again shows that (h) fixes the zero for entropy.

We are now ready to study the relationship between spreads in classical phase space and entropy. We want to find the distribution ρ with zero entropy that minimizes the uncertainty. To do that, we set up a minimization problem using Lagrange multipliers. We want to minimize the product of the variances $\sigma_q^2\sigma_p^2 \equiv \int (q-\mu_q)^2\rho\, dqdp \int (p-\mu_p)^2\rho\, dqdp$, where μ_q and μ_p are the mean position and momentum, under two constraints: ρ integrates to 1 and its entropy is zero. We have:

$$\begin{aligned} L &= \int (q-\mu_q)^2\rho\, dqdp \int (p-\mu_p)^2\rho\, dqdp \\ &\quad + \lambda_1\left(\int \rho dqdp - 1\right) + \lambda_2'\left(-k_B \int \rho \ln(h\rho)\, dqdp - 0\right) \\ &= \int (q-\mu_q)^2\rho\, dqdp \int (p-\mu_p)^2\rho\, dqdp \\ &\quad + \lambda_1\left(\int \rho dqdp - 1\right) + \lambda_2\left(-\int \rho \ln(h\rho)\, dqdp\right) \\ \delta L &= \int \delta\rho[(q-\mu_q)^2\sigma_p^2 + \sigma_q^2(p-\mu_p)^2 + \lambda_1 - \lambda_2 \ln(h\rho) - \lambda_2]dqdp = 0 \\ \lambda_2 \ln(h\rho) &= \lambda_1 - \lambda_2 + (q-\mu_q)^2\sigma_p^2 + \sigma_q^2(p-\mu_p)^2 \\ \rho &= \frac{1}{h} e^{\frac{\lambda_1-\lambda_2}{\lambda_2}} e^{\frac{(q-\mu_q)^2\sigma_p^2}{\lambda_2}} e^{\frac{\sigma_q^2(p-\mu_p)^2}{\lambda_2}} \end{aligned}$$

We recognize the distribution as the product of two independent Gaussians. We can therefore use the standard expression and calculate the entropy, which must be zero. We find:

$$\begin{aligned} \rho &= \frac{1}{2\pi\sigma_q\sigma_p} e^{-\frac{(q-\mu_q)^2}{2\sigma_q^2}} e^{-\frac{(p-\mu_p)^2}{2\sigma_p^2}} \\ I[\rho] &= k_B \ln\left(2\pi e \frac{\sigma_q\sigma_p}{h}\right) = 0 = k_B \ln 1 \\ \sigma_q\sigma_p &= \frac{h}{2\pi e} = \frac{\hbar}{e} \end{aligned}$$

where $\hbar = \frac{h}{2\pi}$. We have found that any distribution $\rho(q,p)$ with non-negative entropy (i.e. that satisfies the third law of thermodynamics) satisfies the inequality

$$\sigma_q \sigma_p \geq \frac{\hbar}{e}. \tag{1.226}$$

Moreover, the inequality is saturated (i.e. becomes an equality) for Gaussian distributions.

Compare this with the Heisenberg uncertainty principle of quantum mechanics, which tells us that the every state satisfies the inequality

$$\sigma_q \sigma_p \geq \frac{\hbar}{2}, \tag{1.227}$$

which is saturated by Gaussian states. Given the parallel, we call equation 1.226 the classical uncertainty principle. Note that the classical uncertainty principle has a clear physical explanation: the third law of thermodynamics. Given that all quantum states have non-negative entropy, the same explanation can be taken for quantum mechanics as well.

Note that, since $e \approx 2.71828 > 2$, the classical uncertainty principle is has a slightly lower bound. The reason is that, at a given entropy, the product $\sigma_q \sigma_p$ is minimized when position and momentum are uncorrelated. Given that the uncorrelated Gaussians are also independent, this means that position and momentum are independent when the uncertainty is the lowest, and the joint distribution is the product of the marginal. This cannot happen in quantum mechanics, since there is cannot write the joint distribution of position and momentum. Classical mechanics, then, can reach a lower uncertainty precisely because position and momentum of the same degree of freedom can be independent variables.

This result also tells us the problem with Maxwell's demon. In this setup, a gas is separated into two chambers with a door. A demon that sees the exact position and momentum of all particles can open the door letting a particle go only from one specific side to the other, decreasing the entropy of the system. But knowing the exact position of all particles corresponds to having access to a state of minus infinite entropy, which is forbidden by the third law. Naturally, if we have access to a source of minus infinite entropy, we can always decrease the entropy of any system by a finite amount without violating the second law: given that minus infinity minus a finite amount is equal to minus infinity, the entropy is technically not decreasing.

This final result should really make it clear that the foundations of statistical mechanics and thermodynamics are not separate from the foundations of classical mechanics and quantum mechanics. In any theory, states are better understood as ensembles, and pure states should be understood as the most precise ensemble that can be prepared in the theory.

1.15 Summary

Let's step back and sum up all that we learned by applying the reverse physics approach to classical mechanics. If we start from scratch, the first assumption that we need to set is Infinitesimal Reducibility IR. This tells us that the domain of classical physics is those objects that can be thought of as being made of arbitrarily small parts. For these objects, specifying the state of the whole system is equivalent to specifying the state of all the infinitesimal parts. In this context, a particle is an infinitesimal part, the limit of recursive reduction of parts into smaller parts. Mathematically, the assumption tells us that the state space of the particles is

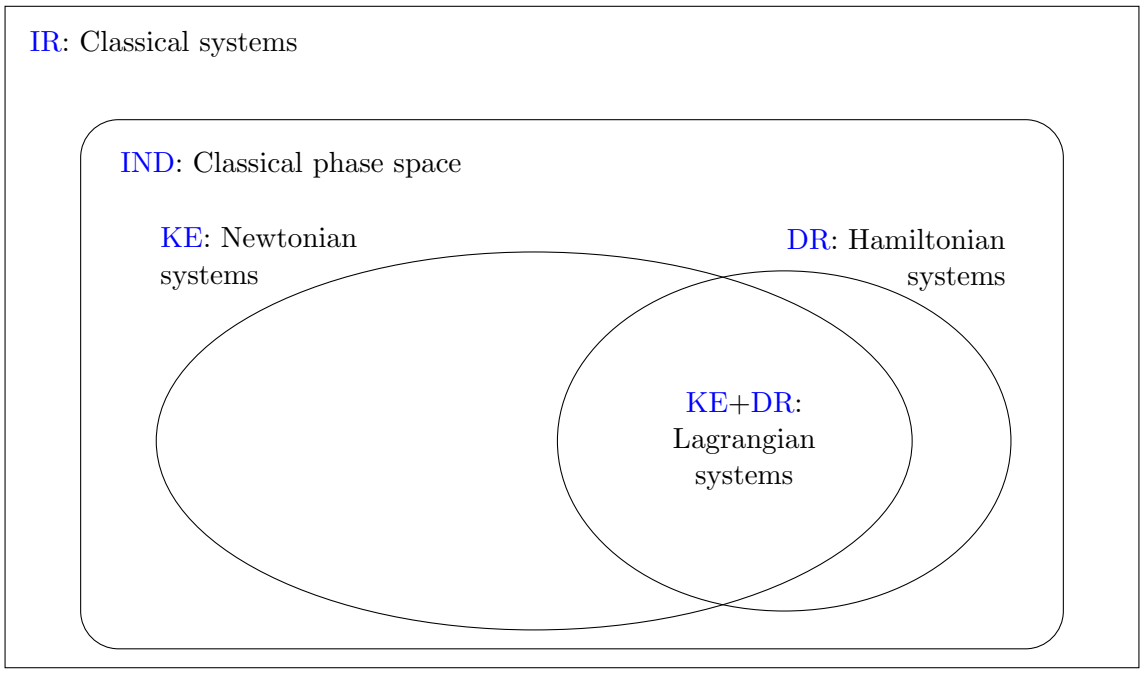


Figure 1.9: Relationship between the constitutive assumptions and the different formulations of classical mechanics. All classical systems satisfy infinitesimal reducibility. The independence of DOF recovers classical phase space. An ideal gas, with state variables $[P\,V\,T]$ is an example of a classical system that does not satisfy IND. Another example is a system with non-holonomic constraints. Newtonian system are fully characterized by spatial trajectories. Hamiltonian systems are characterized by deterministic and reversible dynamics. Lagrangian systems are characterized by both.

a differentiable manifold, together with a volume measure that defines the count of states for each region. The state space of a classical object is a distribution over such manifold.

We then add assumption IND that tells us that the system is decomposable into independent degrees of freedom. This means that not only are we able to count states over regions of state space, but we must be able to count configurations along each DOF. Moreover, the count of configurations over multiple DOFs must be the product of the count over each DOF. The only way to obtain this structure, so that the count is defined independently of the units used to describe states, is for the manifold to be even dimensional, and for any state variable q^i that defines an independent unit, there is a conjugate state variable p_i whose units are count of states over units of the corresponding q^i. This recovers the structure of classical phase space. Mathematically, this gives us the structure of a symplectic manifold, where the symplectic form ω_{ab} counts the configurations over the infinitesimal parallelogram defined by two vectors.

If we add assumption DR that tells us that the evolution is deterministic and reversible, we recover Hamiltonian mechanics. For a system to be deterministic and reversible, in fact, we must map not only each initial state to one and only one final state, but we must map initial to final regions while preserving the state count. Moreover, the notion of independence must be preserved, and therefore the way configurations are counted over independent DOFs must also

be preserved. Under these conditions, the displacement vector field, which tells us how states move in phase space, allows a potential H, which corresponds to the Hamiltonian. This can be generalized to the time-dependent case, which recovers relativistic features even without the notion of a metric tensor. Mathematically, the form ω_{ab} must be preserved, meaning deterministic and reversible evolution is a symplectomorphism. Each symplectomorphism can be characterized by a function H, the Hamiltonian.

If, instead, we add assumption KE that tells us that the dynamics is recoverable from the kinematics, we recover Newtonian mechanics. If the dynamics is recoverable from the kinematics, the momentum must be a function of position and velocity. The units of position are the only independent units, and therefore define the units of momentum and also the transformation between densities over dynamic and kinematic variables. This forces the transformation between position and velocity to be linear. If no forces are present, we can find coordinates such that the linear transformation is simply a multiplication by a constant. That is, $p_i = mv^i$. These are the inertial frames in Cartesian coordinates and m is the inertial mass. In this case, assumption DR applies as well, and, since momentum is constant, the velocity is constant. If a force is present, this can be expressed in terms of position and velocity, recovering $F = ma$.

If we take both assumption DR and KE, Lagrangian mechanics is recovered. The action is the line integral over the potential θ_a of the form $\omega_{ab} = -\partial_a \wedge \theta_b$, and since the vector potential is unphysical, so is the action. The variation of the action is physical as, by Stokes' theorem, it will give the surface integral of ω_{ab}, which is zero if and only if the path is always tangent to the direction of motion, which happens only if the path is an actual evolution of the system.

All the elements of classical mechanics, both physically and mathematically, can be understood just in terms of these assumptions and the concepts that they require.

Part II

Physical Mathematics

Physical mathematics is an approach to the mathematical foundations of physics that seeks to construct mathematical structures strictly from axioms and definitions that can be rigorously justified from physical requirements, instead of simply taking tools developed within mathematics and applying them to physics or physics-inspired problems. Physical mathematics is based on the insight that **when physical objects are mapped to the right mathematical objects, the physical requirements map to the mathematical definitions**.

If our goal is to fully rederive physical theories from physical assumptions, we need to have a precise mapping between physical objects and mathematical ones. Understanding the axioms and definitions of the mathematical tools used in a physical theory, then, is not just "mathematical detail" of no concern to the physicist, but rather the precise stipulation of properties that certain physical objects must have under suitable, possibly simplifying, assumptions. In this sense, there is no "correct" structure in a mathematical sense, because the correct structure is the one suited to the physical problem at hand.

It should be clear that mathematicians are generally ill-equipped to determine whether mathematical structures are physically significant. As David Hilbert stated, "Mathematics is a game played according to certain simple rules with meaningless marks on paper." Regarding mathematical axioms, Bertrand Russell claimed, "It is essential not to discuss whether the first proposition is really true, and not to mention what the anything is, of which it is supposed to be true." Mathematics knows the rules of everything but the meaning of nothing. It is therefore unreasonable to expect that the foundations of mathematics, by themselves, can provide any foundation for physics.

In the same way that elaborate correct mathematical theories stem from minimal correct mathematical theories (not elaborate incorrect mathematical theories); that large living creatures grow from small living creatures (not large dead creatures); sophisticated physically meaningful theories come from simple physically meaningful theories (and not from sophisticated meaningless ones). Meaningfulness, like correctness or aliveness, is not something that can be imposed after the fact. Therefore the only way to develop physically meaningful mathematical structures is to develop them from scratch: we cannot simply take higher level mathematical objects and "sprinkle meaning", an interpretation, on top.

The goal of physical mathematics, then, is to find how to turn physical assumptions into precise mathematical requirements, such that we are guaranteed to know what exactly each mathematical object represents and under which physical conditions.

A new standard for scientific rigor

From the above discussion, it follows that the standard of rigor mathematicians have developed for their field is not sufficient for the purpose of physical mathematics. Mathematics only deals with formal systems, whose starting points are a set of definitions and rules that are taken as is. At that point, correctness of the premise cannot be established, only self-consistency. Therefore mathematics fails to deal with the most delicate and interesting parts of the foundations of physics: the physical assumptions and how they are encoded into the formal framework. We therefore need rules and standards for rigorously handling the informal parts of the framework and, since there are no guidelines for this, we set our own standard.

We call an **axiom** a proposition that brings new objects or new properties of established objects within the formal framework. A **definition**, instead, is a proposition that further characterizes objects and properties already present in the formalism. **An axiom or a def-**

inition is well posed only when it is clear what the objects represent physically and what aspects are captured mathematically. Therefore each axiom and definition is composed of two parts. The first characterizes the objects and properties within the informal system, tells us what they represent physically. The second part, typically preceded by "Formally", characterizes the part that is captured by the formal system. Axioms and definitions are followed by a **justification** when it is necessary to explain why the elements in the informal system must be mapped into the formal system in the way proposed. Some definitions are purely formal and as such do not require justifications. As this argument spans both the formal and informal systems, this cannot be a mathematical proof in the modern sense. In particular, the justification for an axiom must argue why those objects must exist.

The above standard makes sure we have a perfect identification between formal and informal objects. All mathematical symbols correspond to physical objects and all the relevant physical concepts are captured by the math. All subsequent propositions and proofs, then, can be carried out in the formal system, where it is easier to check for consistency and correctness. However, all the proofs can, if needed, be translated into the informal language and given physical meaning.

Chapter 1

Ensemble spaces

In this chapter we aim to develop a general theory of states and processes that is applicable to any physical system. The core concept is that of an ensemble as we have found that ensembles in both classical and quantum mechanics have a very similar structure that can be abstracted and generalized. The goal is to find necessary requirements for ensembles that can serve as basic axioms and then further suitable assumptions to recover the different theories (e.g. classical, quantum or thermodynamics).

The basic premise is that physical theories are primarily about ensembles. At a practical level, most of the time we can only prepare and measure statistical properties as we do not have perfect control over any system (i.e. all measurements are really statistical). The cases where properties can be prepared with one hundred percent reliability can still be understood as ensembles of identical preparations. At a conceptual level, the goal of physics is to write laws that can be repeatedly tested: every time that one prepares a system according to a particular procedure and lets it evolve in particular conditions, he will obtain a particular result. That is, the idea of repeatability of experimental results implicitly assumes that the objects of scientific inquiry are not single instances, but the infinite collections of all reproducible instances. This means that any physical theory, at the very least, will have to provide a mathematical representation for its ensembles.

By ensemble we mean what is usually meant in statistical mechanics: we have a preparation device that follows some known recipe; its output is varied but it is consistently varied (i.e. its statistical properties are well defined); the collection of all possible outputs taken as one object is an ensemble. In classical physics, ensembles are probability distributions over the full description of the system, over classical phase space. In quantum mechanics, ensembles are represented by density matrices and density operators. In the standard approach statistical ensembles are defined on top of the space of "true" physical states (e.g. microstates, pure states, ...). We will proceed in the opposite way: we will start from the ensembles and recover the states as the "most pure" ensembles. There are two main advantages: the first is that we can create a theory that is agnostic about what the fundamental states are, and is therefore general. The second is that this approach is more in line with experimental practice: the experimental data is about statistical ensembles only and the pure states are idealizations that are useful as a mental model or for calculation.

At this point, we have identified three main requirements for ensembles. First, they must be experimentally well defined. This means that there need to be enough experimentally ver-

ifiable statements to fully characterize them. This will impose a topology on the ensemble space. Second, we can always perform statistical mixtures: given two ensembles, we can create a third one by selecting the first or the second according to a certain probability. This will impose a convex structure on the ensemble space. Third, ensembles will need a well-defined entropy which quantifies the variability of the elements within the ensemble. Since the variability cannot decrease when performing a statistical mixture and can only increase up to the variability introduced by the selection, the entropy will have to satisfy certain bounds. From these axioms, many general results can be proven. An ensemble space will be a convex subset of a vector space that will extend over a bounded interval in each direction. The concavity of the entropy will impose a metric over the space, turning it into a geometric space. Each ensemble can be characterized by a subadditive measure, which becomes a probability measure in the classical case. This makes the space of physical theories a lot more constrained than one may imagine at first.

Many problem are still open, as they touch unsettled mathematical questions, particularly in the infinite dimensional case. Therefore this chapter will include conjectures that may or may not be true, which the reader is encouraged to try proving (or disproving).

Note: all logs are assumed to be in base 2.

1.1 Review of standard cases

We will start this chapter by reviewing three cases: discrete classical ensembles, continuous classical ensembles and quantum ensembles. These will be useful both to form an intuition for ensembles and to serve as targets that the whole theory needs to reproduce. We will also go though a series of problematic details and exceptions that we will need to address in the development of the general theory.

Discrete classical ensemble spaces

Definition 1.1. *A **discrete classical ensemble space** is the space of probability distributions over a countable sample space equipped with the Shannon entropy. That is, a discrete classical ensemble space $\mathcal{E}$ is the space of probability distributions over a discrete set X of countably many elements. Each element can thus be identified by a sequence p_i such that $\sum_i p_i = 1$, where p_i is the probability associated to each element $x_i \in X$. The entropy is given by $S(\mathsf{e}) = -\sum_i p_i \log p_i$.*

Finite case

The space of classical distributions over a discrete space corresponds to a simplex. In the finite case, the pure states $X = \{x_i\}_{i=1}^n \subset \mathcal{E}$ are finitely many and each ensemble $\mathsf{e} = \sum_i p_i x_i$ is uniquely identified by a decomposition of pure states. Effectively, each ensemble is a probability distribution over the pure states. Mathematically, each point of the space is a convex combination of the vertices. The simplex has a center point, which corresponds to the maximally mixed state, a uniform distribution over all pure states.

The entropy is given by the Shannon entropy $-\sum_i p_i \log p_i$. This means that the entropy of each pure state is zero and the entropy of the maximally mixed state is $\log n$ where n is the number of pure states. The entropy increases as we go from pure states to the maximally mixed state. The level sets (i.e. the fibers) of the entropy form a series of concentric "shells" that foliates the space.

Note that imposing zero entropy on all pure states is a restrictive condition that does not apply in general. To see this, consider the case where the state is defined by the number of molecules for two substances. This space is the product of two independent variables n_a and n_b. If we have a uniform distribution over N_a cases of n_a and N_b cases of n_b, the total number of cases is $N_a N_b$. Therefore the entropy of the joint state is the sum of the entropy of the marginals. However, if we pair n_a with the total number of molecules $n_{(a+b)}$ we have a problem. The issue is that the variable $n_{(a+b)}$ corresponds to a variable number of joint cases. Therefore the case where the ensemble space is a simplex but the entropy is not the Shannon entropy (i.e. it is the Shannon entropy plus the contributions of entropy from each vertex) is a physically meaningful case that should be possible in the general theory.

Countable case

The countable case is, in some respects, not well defined.

The obvious extension is to include all sequences $\{p_i\} \in [0, 1]$ whose sum converges to one (i.e. the space of all probability measures over a countable discrete space). Since we cannot create a uniform distribution over infinitely many cases, there is no center point, there is no barycenter. Effectively, there is a "hole" in the middle.[1]

However, the space of all probability measures is too large. Note that the entropy is not finite for all $\sum_i p_i = 1$ (i.e. infinite convex combinations).[2] Given that we want the entropy to exist and be finite for all ensembles, this generalization does not seem physically warranted.[3]

Also note that expectation values are not guaranteed to be finite either, and requiring a particular observable to be finite further restricts the space. This restriction may be desirable for another reason: a discrete ensemble space has no notion of the ordering of the pure states. Physically, this would mean that the states with 1, 100, or 1 trillion particles are "equally distant" (i.e. all infinite permutations are allowed). Requiring the expectation of the number of particles to be finite (e.g. $\sum_i N(i) p_i < \infty$) should effectively encode the infinite ordering in the rate of convergence of the probability distributions (i.e. not all infinite permutations would be allowed).

Another problem is identifying the correct topology for the space. The entropy defines a notion of orthogonality, as we will see later, based on the disjoint support of distributions. This suggests that we need a space with an inner product, therefore we should require the probability distribution to be square integrable. However, the inner product could be defined on the square root of the probability distribution, more in line with the quantum case, which would be well defined as probability is never negative. It is not yet clear which is the correct case.

No uncountable case

The uncountably infinite case is not physically relevant, as the space cannot be given a second countable discrete topology. Also note that any set of real numbers whose sum is finite can have only countably many non-zero elements. To understand why, note that there can only

[1] It may be useful to characterize this "hole" and the limit points. There should be at least one limit point for each sequence $\{p_i\}$ whose sum converges to a finite $p < 1$. Intuitively, we can keep that part of the distribution constant while we spread the rest uniformly to all other cases. Each should reach a different limit point.

[2] For details, see J. Stat. Mech. (2013) P04010.

[3] This generalization is called a superconvex space in some literature.

be finitely many terms above any particular positive value if their sum is to remain finite. Effectively, the uncountable case would be stitching together infinitely many countable cases.

Continuous classical ensemble spaces

Definition 1.2. *A **continuous classical ensemble space** is the space of probability distributions over classical phase space equipped with the Shannon/Gibbs entropy. That is, it is the space of probability measures $\mathcal{E}$ over a symplectic manifold X that is absolutely continuous with respect to the Liouville measure μ. The entropy is given by the Shannon/Gibbs entropy calculated using the probability density (i.e. the Radon-Nikodym derivative between the probability measure p and the Liouville measure μ). That is, $S(\rho) = -\int_X \rho \log \rho d\mu$ where $\rho = \frac{dp}{d\mu}$.*

In the continuous case, the space of ensembles can be understood as the space of non-negative integrable functions over a symplectic manifold (e.g. over phase space) that integrate to one. That is, if X is a symplectic manifold, then $\mathcal{E} = \{\rho \in L^1(X) \,|\, \rho(x) \geq 0, \int_X \rho(x) d\mu = 1\}$ where $\mu(U) = \int_U \omega^n$ is the Liouville measure. This is a convex set, whose extreme points would be, in the limit, the Dirac measures (i.e. the probability measure all concentrated at a single phase space point). As we cannot reliably prepare a system at an infinitely precise position and momentum, these distributions are not physical. Also, they would correspond to minus infinite entropy. The Dirac measure, and in general all distributions over a set of measure zero, are excluded because they are not absolutely continuous. The absence of the extreme points is important as, when developing standard constructions in the ensemble space, one cannot rely on the existence of extreme points. In general, this points to a difference between the spectra (i.e. the possible values of a random variable) and pure states (i.e. extreme points), which will be even more pronounced in the quantum case.

The symplectic nature of the manifold is required to assign a frame-invariant density to states and a frame-invariant notion of independence between DOFs, as we saw in the classical mechanics section of reverse physics. The entropy is given by $-\int \rho \log \rho d\mu$ where μ is the Liouville measure and ρ is the probability density over canonical coordinates. If a different measure is used, or if the coordinates are not canonical, the formula gives the wrong result.[4]

Similarly to the countable discrete case, the entropy can be infinite and expectation values can be infinite. The added complication is the frame invariance: it would not make sense to have finite expectation for position in one frame but infinite in another. Requiring all functions of position and momentum to have finite expectation restricts the distributions to those with finite support. Requiring all polynomial functions of position and momentum to have finite expectation restricts the distributions to those that decay faster than any polynomial. Note that the expectation of all polynomials of position and momentum are not enough to reconstruct the distribution. Furthermore, derivatives of the distribution over position and momentum are needed to determine how the distribution evolves over time given a time evolution map. This may suggest that the proper space of probability measures does not

[4] It may be interesting to study the shell of zero entropy states. For example, it should not be path connected. All uniform distributions with support of the same finite size (in terms of the Liouville measure) will have the same entropy. The region, however, need not be contiguous. Since we cannot continuously transform a single region into two disjoint regions, there will be different distributions at zero entropy that cannot be transformed continuously.

include all the absolutely continuous ones, but only those for which the probability density is a Schwartz function. These details are still to be understood.

The above consideration would seem to rule out probability measures with a discontinuous probability density. This is somewhat of an unclear point. Originally, we thought the probability density should be continuous as only continuous functions can physically represent experimental relationships. However, the relationship given by the measure is not between points and probability but rather between sets and probability. The probability density is, in a sense, not the prime physical object. The measure is. The probability density, in fact, is not uniquely defined as countably many discontinuities can be added and the measure is not changed. This seems to suggest that what happens at a single point, or over a set of measure zero, is not critical. However, it is unclear how to recover continuity of the space without continuity of the distributions. It could be that it is the space of deterministic and reversible transformations that defines continuity. One added benefit of allowing discontinuous probability densities is that the uniform distribution gives the maximum entropy within the set of probability distribution with support over a finite measure set. This makes entropy maximization considerations a lot easier.

Unlike the discrete classical case, subspaces and dimensionality of subspaces cannot be defined without the entropy. The issue is that we need a measure on the set of pure states, and the convex structure cannot provide it. The entropy does, however, as the supremum of the entropy for all distributions with support U is $\log \mu(U)$. As we will see, the entropy can be used to both identify subspaces and recover the Liouville measure.

Quantum ensemble spaces

Definition 1.3. *A **quantum ensemble space** is the space given by the density matrices/operators of a Hilbert space equipped with the von Neumann entropy. That is, given a separable Hilbert space $\mathcal{H}$ for a quantum system, the ensemble space $\mathcal{E}$ is the space of positive semi-definite self-adjoint operators with trace one $M(\mathcal{H})$. The space of pure states X is given by the projective space $P(\mathcal{H})$. The entropy of an ensemble $\rho \in \mathcal{E}$ is given by the von Neumann entropy $S(\rho) = -\operatorname{tr}(\rho \log \rho)$.*

Finite dimensional case

The simplest non-trivial case is the qubit, for which the Bloch ball is the space of ensembles $\mathcal{E} = M(\mathcal{H})$. The interior of the Bloch ball corresponds to mixtures while the surface corresponds to the pure states $X = P(\mathcal{H}) = \{|\psi\rangle\langle\psi|\}_{\psi\in\mathcal{H}}$. In quantum ensemble spaces there is no unique decomposition in terms of pure states. Note that the space is exactly characterized by knowing which different mixtures provide the same ensemble.

Multiple decompositions make the ensemble space behave in a way that is a hybrid between the classical discrete and continuous. Pure states are properly a part of the ensemble space, as in the discrete case, and we can describe each mixture in terms of finitely many pure states. However, the pure states form a continuum, therefore we can also define probability densities over the space, convex integrals. For example, for a single qubit, the maximally mixed state (the center of the ball) can be equally described as the equal mixture of any two opposite states (e.g. spin up and spin down, or spin left and spin right). However, it can also be described as the equal mixture of the whole sphere.

Note that complex projective spaces are symplectic, which is what allows one to define frame invariant densities. The goal is to have one argument applied to the generic definition

as to why the space of pure states must be symplectic. Also note that the two dimensional sphere is the only symplectic sphere. By homogeneity, we should be able to argue that the space is symmetric around the maximally mixed state, and is therefore a sphere. The symplectic requirement would select dimension two. Note that real and quaternionic spaces would be excluded by this argument.

The von Neumann entropy for the maximally mixed state is $\log n$ where n is the dimensionality of the Hilbert space. Again we see that the maximum entropy gives us a measure of the size of the space. Note that, to calculate the von Neumann entropy, we are diagonalizing the density matrix ρ. This means finding a set of orthogonal pure states x_i such that $\rho_i = \sum_i p_i x_i$ is a convex combination. Note that the convex hull of a set of n orthogonal pure states is an n-dimensional simplex whose center is the maximally mixed state. Therefore, we are looking for a simplex that contains ρ and the maximally mixed state. In the two dimensional case, ρ is an interior point of the Bloch ball. Take the line that connects ρ to the center. The two points of the sphere are the extreme points for the decomposition. The distance from the points will be proportional to the probability. Because of this property, the von Neumann entropy is the smallest Shannon entropy among all possible decompositions.

Countably infinite dimensional case

The countably infinite dimensional case presents similar problems as the classical case, and adds others. As in the classical infinite cases, the maximally mixed state (i.e. uniform distribution) is not in the convex space and the entropy is not finite for all infinite convex combinations. As in the classical continuous case, there is the issue of finite expectation of position/momentum in all frames. The problem is compounded by the fact that one cannot require finite expectation for all functions of position and momentum: finite support in position automatically implies infinite support on momentum, since the distribution in momentum is the Fourier transform of that in position.

The Hilbert space for a discrete variable with infinite range (e.g. number of particles) and a continuous variable (e.g. position/momentum) is the same. The first is defined as the space of square-summable complex sequences l^2 while the second is the space of square integrable complex functions L^2. Given that L^2 allows a countable basis, the two are isomorphic. This also means that all spaces with finitely many degrees of freedom are also isomorphic. This makes the problem of infinite expectations even more problematic.

Note that Schwartz spaces have finite expectation for all polynomial functions of position and momentum, given that the momentum operator is the derivative of position. Given that infinite permutations can change the rate of convergence, the Schwartz space has an idea of what is further away from the origin, unlike Hilbert spaces. We will likely want to use Schwartz spaces instead of Hilbert spaces to make the physics and mathematics more consistent.

1.2 Axiom of ensemble and topology

Statistical ensembles will be the cornerstone of our general theory for states and processes. In this section we will see how any physical theory must, at least, define a space of ensembles which define the output of all possible processes considered by the theory. Since a physical theory must allow for experimental verifiability, the ensemble space must be endowed with a T_0 second countable topology.

We saw how the principle of scientific objectivity required science to be universal, non-contradictory and evidence based. If our goal is to find laws that govern the evolution of physical systems, however, this is not sufficient. Scientific laws will be statements of the type *"every time we prepare this type of system according to this procedure and let it evolve under these conditions, we will find the system in this configuration after some time."* "Every time" implies the principle of scientific reproducibility.

Principle of scientific reproducibility. Scientific laws describe relationships that can always be experimentally reproduced.

Consider the hypothesis that all life on earth descends from a single common ancestor. This is a scientific hypothesis that may be experimentally falsified, but it is about a single event. As such, it is not a scientific law. The theory of evolution through natural selection, instead, describes what always happens to a population given a set of circumstances, and is therefore a law. As such, it does not describe a particular set of living organisms or traits, it applies to all of them but, as a consequence, to none in particular.

The same applies to the laws of physics. Classical Hamiltonian mechanics or quantum mechanics will apply to certain classes of physical systems, describing the common behavior within each class over all possible instantiations, but none in particular. That is, the law is not describing a particular behavior of a particular system, but the common behavior of the aggregate of all similarly prepared systems at all possible times.

The subject of a physical law, then, is not a single system in a single state, but an ensemble: all possible preparations of equivalent systems prepared according to the same procedure. A general theory of states and processes, then, will be a theory about ensembles as this is the least restrictive requirement needed. Any physical theory will *at least* provide us with a set of ensembles, and the physical laws must be able to describe the evolution of those ensembles.

Given that we are still talking about scientific investigation, ensembles must be experimentally well defined and the principle of scientific objectivity applies. This means that ensembles are the possibilities of an experimental domain, which means points of a T_0 second countable topological space where the topology corresponds to the natural one defined by the verifiable statements.

The verifiable statements for the ensemble space are statements about the ensembles themselves, either in terms of statistical quantities (e.g. *"the average energy of the particle is* $3 \pm 0.5\, eV$*"*) or in terms of preparation settings (e.g. *"the beam goes through a polarizer oriented vertically within 1 degree"*). Probability ranges are also typical verifiable statements on ensembles (e.g. *"the coin toss will result in heads between 49 and 51 percent of the cases"*). Note that the verifiable statements at the level of the ensemble are different from the verifiable statements at the level of each instance. Saying that a coin is fair, for example, is a statement on the ensemble while saying that the outcome at a particular time was heads is a statement on the instance. The two are unrelated: whether the coin is fair is an independent statement with respect to a particular instance. Therefore the topology of the ensemble space and the topology of the random variables are distinct conceptual and mathematical objects (e.g. the topology of the Bloch ball is the standard topology of $\mathbb{R}^3$ but the topology on the values of spin along a given direction is the discrete topology), and it will be much later that the two will be reconciled.[5]

[5] In fact, many details are still open.

It should also be clear that ensembles are theoretical objects, idealizations: an infinite collection of instances cannot be realized. What is realized in a laboratory will be a finite version, with all limitations in terms of precision that go with it. One may ask: how can something so idealized represent physical objects? But this is exactly what we do in all other areas of physics: we talk about spheres, perfect fluids, isolated systems or immovable objects. All of these are idealized objects, and we model the world with such abstractions. Ensembles are useful idealizations precisely because they ignore details that are not relevant for the problem at hand. If we want to write physical laws, in fact, we can only write them on those features that are common to all instances. The ensemble represents exactly those and only those features.

Setting ensembles as a primary notion also solves another conceptual problem. The ensemble is not constructed as a limit of infinite instances, which would pose a number of problems. The ensemble describes the preparation procedure, and therefore the collection of instances is potential and comes before the instances. For example, a fair coin can be understood as the collection of instructions for producing and throwing a fair coin. The fact that a fair coin will produce, in a large trial, about half heads and half tails is a consequence of the type of preparation. This gives automatically an interpretation of probability that is more along the lines of propensity, which is more appropriate to express objective causal relations.

Axiom 1.4 (Axiom of ensemble)**.** *The state of a system is given by an* ***ensemble****, which represents the collection of all possible outputs of a preparation procedure for a physical system. The set of all possible ensembles for a physical system is its* ***ensemble space****. Formally, an ensemble space is a* T_0 *second countable topological space where each element is called an ensemble.*

Justification. In experimental settings, preparation procedures never prepare a system exactly in the same configuration. Experimental results, then, are always in terms of statistical preparations and statistical measurements. A physical theory must be able to talk about the possible statistical descriptions within the theory. States, then, can be understood as ensembles, idealized statistical descriptions, as those are what is connected to experimental practice.

Equivalently, reproducibility is a basic requirement of a physical theory. A physical law, then, must be understood as describing a relationship that always exists whenever the same set of circumstances is replicated. Given that we need to always be able to replicate those circumstances "one more time", the relationship is about countably infinite preparations and results: ensembles. Therefore, to the extent that physics is about reproducible experimental results, the basic theoretical description of a system is in terms of ensembles. This justifies the use of ensembles as the fundamental object to describe the state of a system.[a]

Ensembles are experimentally defined objects, and therefore they are possibilities of an experimental domain. This means that an ensemble space is a T_0 second countable topological space where each element is an ensemble and the topology is induced by the verifiable statements. □

[a] Note that reproducibility also already implies that all properties that characterize an ensemble must be relative to the procedure. If the properties depended, for example, on absolute space or absolute time, then different practitioners would not be able to prepare the same ensemble.

We should now verify that our three standard cases satisfy the axiom of ensemble.

Proposition 1.5. *Classical discrete, classical continuous and quantum ensemble spaces satisfy the axiom of ensemble.*

Proof. For the classical discrete case, including all infinite convex combinations, we have the subset of ℓ^1 of all non-negative sequences that sum to one. Similarly, for the classical continuous case we have the subset of L^1 that corresponds to non-negative distributions that integrate to one. Both of these spaces, with the subspace topology, are separable, admit a countable orthonormal basis and can be given a topology that is T_0 and second countable. Note that, in case the correct formulation is in terms of square integrable functions, instead of simply integrable functions, the space is still T_0 and second countable.

For the quantum case, the Hilbert space with its standard topology is also separable, will admit a countable orthonormal basis and can be given a topology that is T_0 and second countable. A density operator will be fully defined by its result on the basis, which means the space of ensembles is also a vector space with a countable basis and can be given a topology that is T_0 and second countable. □

1.3 Axiom of mixture and convex structure

In this section we are going to see how the ability to perform statistical mixtures leads to a convex structure for ensemble spaces. Only mixtures of finitely many elements are guaranteed to exist, and the topology will tell us which infinite mixtures are possible. The convex structure also gives us a basic notion to compare ensembles: one ensemble can be a component of another if the second can be seen as a mixture of the first with something else. Two ensembles are separate if they have no common component.

As we saw before, an ensemble is the collection of all outputs of a preparation procedure. The idea is that we can always combine preparation procedures selecting the output among them with a given probability distribution. This statistical mixture is another preparation procedure which will correspond to an ensemble. The ensemble space of any physical theory, then, must allow statistical mixtures, which leads to a convex structure.

Definition 1.6. *Given a real number $p \in [0,1]$, its complement is defined as $\bar{p} = 1 - p$.*

Axiom 1.7 (Axiom of mixture)**.** *The statistical mixture of two ensembles is an ensemble. Formally, an ensemble space $\mathcal{E}$ is equipped with an operation $+ : [0,1] \times \mathcal{E} \times \mathcal{E} \to \mathcal{E}$ called **mixing**, noted with the infix notation $p\mathsf{a} + \bar{p}\mathsf{b}$, with the following properties:*

- ***Continuity**: the map $+(p, \mathsf{a}, \mathsf{b}) \to p\mathsf{a} + \bar{p}\mathsf{b}$ is continuous (with respect to the product topology of $[0,1] \times \mathcal{E} \times \mathcal{E}$)*
- ***Identity**: $1\mathsf{a} + 0\mathsf{b} = \mathsf{a}$*
- ***Idempotence**: $p\mathsf{a} + \bar{p}\mathsf{a} = \mathsf{a}$ for all $p \in [0,1]$*
- ***Commutativity**: $p\mathsf{a} + \bar{p}\mathsf{b} = \bar{p}\mathsf{b} + p\mathsf{a}$ for all $p \in [0,1]$*
- ***Associativity**: $p_1\mathsf{e}_1 + \bar{p}_1\left(\frac{p_2}{\bar{p}_1}\mathsf{e}_2 + \frac{p_3}{\bar{p}_1}\mathsf{e}_3\right) = \bar{p}_3\left(\frac{p_1}{\bar{p}_3}\mathsf{e}_1 + \frac{p_2}{\bar{p}_3}\mathsf{e}_2\right) + p_3\mathsf{e}_3$ where $p_1, p_3 \in [0,1)$ and $p_1 + p_3 \leq 1$ and $p_2 = 1 - p_1 - p_3$*

Justification. This axiom captures the ability to create a mixture merely by selecting between the output of different processes. Let e_1 and e_2 be two ensembles that represent the output of two different processes P_1 and P_2. Let a selector S_p be a process that outputs two symbols, the first with probability p and the second with probability $\bar{p}$. Then we can create another process P that, depending on the selector, outputs either the output of P_1 or P_2. All possible preparations of such a procedure will form an ensemble. Therefore we are justified in equipping an ensemble space with a mixing operation that takes a real number from zero to one, and two ensembles.

Given that mixing represents an experimental relationship, and all experimental relationships must be continuous in the natural topology, mixing must be a continuous function. In general, the mixing coefficient p corresponds to a value from a continuously ordered quantity between zero and one, as defined in the previous chapter, and therefore the natural topology is the one of the reals.[a] This justifies continuity.

If $p = 1$, the output of P will always be the output of P_1. This justifies the identity property. If P_1 and P_2 are the same process, then the output of P will always be the output of P_1. This justifies the idempotence property. The order in which the processes are given does not matter as long as the same probability is matched to the same process. The process P is identical under permutation of P_1 and P_2. This justifies commutativity. If we are mixing three processes P_1, P_2 and P_3, as long as the final probabilities are the same, it does not matter if we mix P_1 and P_2 first or P_2 and P_3. This justifies associativity. □

Corollary 1.8. *An ensemble space is a convex space.*

Proof. The properties of the axiom of mixture match the basic definition of convex spaces. For example, see https://ncatlab.org/nlab/show/convex+space or arXiv:0903.5522. The notation and terminology will be slightly different to better map to physics ideas. □

[a] It may be argued that rational numbers could be prepared exactly, as one may design a procedure that, for example, alternates the selection deterministically. Therefore one could have a countable subset of topologically isolated ensembles. It is not clear whether this would create problems or not. Since the topology of the reals would still be required as a subspace topology anyway, we leave investigating this case to future work.

As we progress through the details of the theory, we will see that all linear structures in physics are, in one way or another, manifestations of this basic structure. For example, the linearity of the Hilbert space in quantum mechanics is connected to the linearity of density operators and expectations.

Before proceeding, we should now check that the axiom of mixture is satisfied by the standard cases.

Proposition 1.9. *Discrete classical ensemble spaces, continuous classical ensemble spaces and quantum ensemble spaces satisfy the axiom of mixture.*

Proof. The space $\mathcal{E}$ of probability measures, discrete or continuous, is a convex subset of the topological vector space of signed finite measures. It is therefore closed under convex combinations: if $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ are probability measures, then $p\mathsf{a} + \bar{p}\mathsf{b}$ is a probability measure. The properties of mixing are inherited from the properties of linear combinations, which include continuity. Therefore the discrete and continuous classical ensemble spaces satisfy

the axiom of mixture.

Similarly, the space of positive semi-definite self-adjoint operators with trace one is a convex subset of the topological vector space of self-adjoint operators. Therefore it is closed under convex combinations, which are continuous in the given topology, and it will satisfy the axiom of mixture. □

Finite and infinite mixtures

The axiom of mixture only guarantees the existence of a mixture between two ensembles. We can mix elements recursively and extend the operation to finitely many ensembles. Commutativity and associativity make these mixtures independent of the mixing order, such that they depend only on the ensembles chosen and the mixing coefficients associated to each ensemble.

Definition 1.10 (Finite mixture)**.** *Let* $\{e_i\}_{i=1}^n \subseteq \mathcal{E}$ *be a finite subset of ensembles and* $\{p_i\} \in [0,1]$ *be a finite set of coefficients such that* $\sum_{i=1}^n p_i = 1$*, then the* ***finite mixture****, noted* $\mathsf{a} = \sum_{i=1}^n p_i \mathsf{e}_i$*, is defined to be*

$$p_1\mathsf{e}_1 + (1-p_1)\left(\tfrac{p_2}{1-p_1}\mathsf{e}_2 + \tfrac{1-p_1-p_2}{1-p_1}\left(\tfrac{p_3}{1-p_1-p_2}\mathsf{e}_3 + \tfrac{1-\sum_{i=1}^3 p_i}{1-p_1-p_2}\Big(\dots + \tfrac{1-\sum_{i=1}^{n-2} p_i}{1-\sum_{i=1}^{n-3} p_i}\left(\tfrac{p_{n-1}}{1-\sum_{i=1}^{n-2} p_i}\mathsf{e}_{n-1} + \tfrac{p_n}{1-\sum_{i=1}^{n-2} p_i}\mathsf{e}_n\right)\Big)\Big)\right)\right). \tag{1.11}$$

Consistency check. For the definition to work, we need to make sure that the final ensemble does not depend on the order of mixing. We first check with three elements. Let $p_1, p_2, p_3 \in [0,1]$ with $p_1 + p_2 + p_3 = 1$. By commutativity, we can switch the second with the third element:

$$\begin{aligned} p_1\mathsf{e}_1 + p_2\mathsf{e}_2 + p_3\mathsf{e}_3 &= p_1\mathsf{e}_1 + \bar{p}_1\left(\frac{p_2}{\bar{p}_1}\mathsf{e}_2 + \frac{p_3}{\bar{p}_1}\mathsf{e}_3\right) = p_1\mathsf{e}_1 + \bar{p}_1\left(\frac{p_3}{\bar{p}_1}\mathsf{e}_3 + \frac{p_2}{\bar{p}_1}\mathsf{e}_2\right) \\ &= p_1\mathsf{e}_1 + p_3\mathsf{e}_3 + p_2\mathsf{e}_2. \end{aligned} \tag{1.12}$$

Then, by commutativity and associativity, we can switch the first with the third element:

$$\begin{aligned} p_1\mathsf{e}_1 + p_2\mathsf{e}_2 + p_3\mathsf{e}_3 &= p_1\mathsf{e}_1 + \bar{p}_1\left(\frac{p_2}{\bar{p}_1}\mathsf{e}_2 + \frac{p_3}{\bar{p}_1}\mathsf{e}_3\right) = p_1\mathsf{e}_1 + \bar{p}_1\left(\frac{1-p_1-p_3}{\bar{p}_1}\mathsf{e}_2 + \frac{p_3}{\bar{p}_1}\mathsf{e}_3\right) \\ &= \bar{p}_3\left(\frac{p_1}{\bar{p}_3}\mathsf{e}_1 + \frac{1-p_1-p_3}{\bar{p}_3}\mathsf{e}_2\right) + p_3\mathsf{e}_3 = p_3\mathsf{e}_3 + \bar{p}_3\left(\frac{p_1}{\bar{p}_3}\mathsf{e}_1 + \frac{p_2}{\bar{p}_3}\mathsf{e}_2\right) \\ &= p_3\mathsf{e}_3 + p_2\mathsf{e}_2 + p_1\mathsf{e}_1. \end{aligned} \tag{1.13}$$

Since switching the first with the second is equivalent to switching the second with the third and the third with the first, we can reach all permutations.

Note that the definition is recursive, therefore we can use proof by induction. The base case is a sequence of two elements. By commutativity, the order does not matter. Given a sequence of n elements, the inductive hypothesis is that the order does not matter for the last $n-1$ elements. Therefore, it suffices to show that we can switch the first and the second element. Note that we can sum the last $n-2$ elements, thus converting this to a

problem of three elements. We proved before that we can switch the first with the second element, and we can now re-expand the third element into the full sequence. Therefore, the order of mixing does not matter for any finite mixture. □

Remark. Note that we can collect and expand convex combinations into other convex combinations. Because coefficients always sum to one, when breaking the expression in two, we can always calculate the new coefficients from one part. For example, $p_a\mathsf{a} + p_b\mathsf{b} + p_c\mathsf{c} + p_d\mathsf{d} = p(\frac{p_a}{p}\mathsf{a} + \frac{p_b}{p}\mathsf{b}) + \bar{p}(\frac{p_c}{\bar{p}}\mathsf{c} + \frac{p_d}{\bar{p}}\mathsf{d})$ where $p = p_a + p_b$ is defined only on the left part. Since $1 = p + \bar{p} = p_a + p_b + p_c + p_d$, we automatically have that $\bar{p} = p_c + p_d$.

While mixtures of finitely many elements are always guaranteed to exist, the extension to mixture of infinite elements is not. First of all, an infinite mixture of ensembles with finite entropy does not necessarily have finite entropy.[6] Secondly, it may lead to infinite expectation values which make ensembles not physically meaningful. For example, suppose we put i particles in a box according to the distribution $\frac{6}{\pi^2 i^2}$ given that $\sum_{i=1}^{\infty} \frac{6}{\pi^2 i^2} = 1$. The expectation $\sum_{i=1}^{\infty} \frac{6}{\pi^2 i^2} i$ diverges. Given that every finite preparation will have a finite expectation value, the actual implementation of that ensemble will necessarily give us a stream of preparations whose number of particles must keep increasing. No finite statistics is representative of the ensemble and, worst of all, the finite statistics will necessarily have averages of arbitrarily large differences. We therefore cannot simply look at the coefficients to understand whether the infinite mixture gives us a valid ensemble or not.

Whether an infinite mixture $\sum_{i=1}^{\infty} p_i\mathsf{e}_i$ converges or not in the ensemble space is therefore determined by the topology (i.e. experimental verifiability). The axiom guarantees only finite mixtures and we let the closure of the topology handle the limits.

Definition 1.14 (Infinite mixture)**.** *Let $\{e_i\}_{i=1}^{\infty} \subseteq \mathcal{E}$ be a sequence of ensembles and $\{p_i\} \in [0,1]$ be a sequence of coefficients such that $\sum_{i=1}^{\infty} p_i = 1$. Then the ensemble a is an **infinite mixture** of those ensembles if it is a topological limit of the sequence of finite mixtures $\sum_{i=1}^{n} \frac{p_i}{p_n}\mathsf{e}_i$, where $p_n = \sum_{i=1}^{n} p_i$. If the infinite mixture is unique, we write $\mathsf{a} = \sum_{i=1}^{\infty} p_i\mathsf{e}_i$.*

Remark. We will see later that the entropy constrains the topology to be Hausdorff, therefore all infinite mixtures, if they exist, are unique.

It is unclear whether commutativity and associativity extend, or should extend, to infinite mixtures. For series, unconditional convergence defines convergence that does not depend on infinite reordering, but it is unclear even whether this is a desirable property. Additionally, we would expect that if an infinite mixture is possible, any submixture should converge as well. That is, if $\sum_{i=1}^{\infty} p_i\mathsf{e}_i$ converges, then $\sum_{i\in I} \frac{p_i}{p_I}\mathsf{e}_i$ converges for all $I \subseteq \mathbb{N}$ where $p_I = \sum_{i\in I} p_i$.

We do know that the convex structure is not enough to guarantee the above property, as shown by the following counterexample provided on stack exchange. Consider $\mathbb{R}$. It is a metrizable second-countable topological vector space, which means it is a convex space with a topology such that the mixing operation is continuous. It satisfies the axiom of ensemble and the axiom of mixture. Let $p_i = \frac{6}{\pi^2 i^2}$, which means $\sum_i p_i = 1$ and $\mathsf{e}_i = (-1)^i i$. We have

$$\sum_{i=1}^{\infty} p_i\mathsf{e}_i = \frac{6}{\pi^2}\sum_{i=1}^{\infty} \frac{(-1)^i}{i} = -\frac{6}{\pi^2}\ln 2$$

[6] For details, see J. Stat. Mech. (2013) P04010.

which means it converges. However, let $I = \{2, 4, 6, \dots\}$. We have

$$p_I = \sum_{k=1}^{\infty} p_{2k} = \sum_{k=1}^{\infty} \frac{6}{\pi^2 4k^2} = \frac{1}{4}$$
$$\sum_{i\in I} \frac{p_i}{p_I} \mathsf{e}_i = 4 \sum_{i\in I} p_i \mathsf{e}_i = 4 \sum_{k=1}^{\infty} \frac{6}{\pi^2 (4k^2)} 2k = \frac{12}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k} \to \infty.$$

Therefore the convex combination of all even elements of the series diverges.

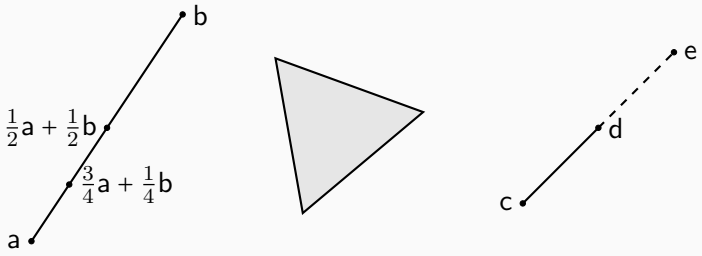
Note that the above counterexample works because the series is effectively the sum of two divergent series. The axiom of entropy will force the ensemble space to have a bounded intersection with every affine line. Therefore the above problem would not exist, because one cannot produce a divergent series from a bounded interval of the real line. It is not clear, though, whether this is enough to show that submixtures of convergent infinite mixtures converge. We leave this as conjecture 1.113.

TODO: add conjecture for definition of convex integrals

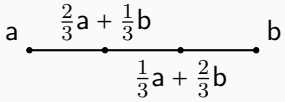
Common components and separateness

The convex structure allows us to characterize ensembles based on whether they can be mixed into one another. For example, we can ask whether an ensemble is or is not the mixture of some other ensembles; or whether two ensembles can be expressed as a mixture of a common component.

Definition 1.15. *Let $\mathcal{E}$ be an ensemble space. Let $\mathsf{a} = \sum_i p_i \mathsf{e}_i$ where $\mathsf{a}, \mathsf{e}_i \in \mathcal{E}$ and $p_i \in (0, 1]$ such that $\sum_i p_i = 1$. We say that a is a **mixture** of $\{\mathsf{e}_i\}$, each e_i is a **component** of a and each p_i is a **mixture coefficient**.*



Remark. In terms of the convex space, all the mixtures of two ensembles correspond to the segment between them; all the mixtures of three ensembles correspond to the triangle formed by the three elements and so on. An ensemble c is a component of a different ensemble d if the segment connecting c and d can be extended past d. If two elements are not a component of each other, then they are the extreme points of the line that goes through the two. That is, the segment cannot be extended.



Remark. Note that two ensembles can be components of each other. Consider $\frac{2}{3}\mathsf{a} + \frac{1}{3}\mathsf{b}$ and $\frac{1}{3}\mathsf{a} + \frac{2}{3}\mathsf{b}$. We can write $\frac{2}{3}\mathsf{a} + \frac{1}{3}\mathsf{b} = \frac{1}{2}\left(\frac{1}{3}\mathsf{a} + \frac{2}{3}\mathsf{b}\right) + \frac{1}{2}\mathsf{a}$ and $\frac{1}{3}\mathsf{a} + \frac{2}{3}\mathsf{b} = \frac{1}{2}\left(\frac{2}{3}\mathsf{a} + \frac{1}{3}\mathsf{b}\right) + \frac{1}{2}\mathsf{b}$ (they are both midpoints along each other). Therefore a component is not necessarily "smaller" or "better defined" than the mixture. Mathematically, "being a component of" is not a partial order. It is reflexive and transitive, but it is not antisymmetric. In practical terms, we need something else to tell us whether we are, for example, taking a limit with components that become "smaller and smaller."

We can also characterize some ensembles based on what other ensembles they can admit as components. An extreme point is an ensemble that has only itself as a component. For example, a pure state will be an extreme point as it cannot be expressed as a mixture of any other states. Conversely, an internal point is an ensemble that admits any other ensemble as a component. For example, in a finite discrete classical space, the uniform distribution over all cases can be seen as the mixture of any other distribution with something else. A boundary point is an ensemble that is not an internal point. The figure helps visualize the properties.

Definition 1.16. *Let $\mathcal{E}$ be an ensemble space. An* ***extreme point*** $\mathsf{e} \in \mathcal{E}$ *is an ensemble that has no component distinct from itself. That is, there is no* $\mathsf{a} \in \mathcal{E} \setminus \{\mathsf{e}\}$ *such that* $\mathsf{e} = p\mathsf{a} + \bar{p}\mathsf{b}$ *for some* $p \in (0, 1]$ *and* $\mathsf{b} \in \mathcal{E}$*. An* ***internal point*** $\mathsf{e} \in \mathcal{E}$ *is an ensemble for which every ensemble is a component. That is, for every* $\mathsf{a} \in \mathcal{E}$ *there is always* $\mathsf{b} \in \mathcal{E}$ *and* $p \in (0, 1]$ *such that* $\mathsf{e} = p\mathsf{a} + \bar{p}\mathsf{b}$*. A* ***boundary point*** *is any ensemble that is not an internal point.*

Remark. The notion of internal point parallels the similar notion in topological vector spaces.

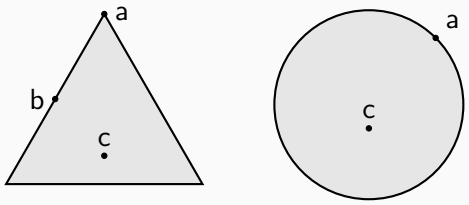


Figure 1.1: In both spaces, a is an extreme point while c is an internal point; b is a boundary point, but is not an extreme point (it is the mixture of the vertices on the same side). On the circle, all boundary points are extreme points.

Proposition 1.17. *The set of all internal points $I_{\mathcal{E}}$ is a convex set.*

Proof. Let $\mathsf{e}_1, \mathsf{e}_2 \in I_{\mathcal{E}}$ be two internal points of $\mathcal{E}$. Since e_1 and e_2 are internal points, given any $\mathsf{a} \in \mathcal{E}$ we can find $\mathsf{b}_1, \mathsf{b}_2 \in \mathcal{E}$ such that

$$\begin{aligned} \mathsf{e}_1 &= p_1\mathsf{a} + \bar{p}_1\mathsf{b}_1 \\ \mathsf{e}_2 &= p_2\mathsf{a} + \bar{p}_2\mathsf{b}_2 \end{aligned} \tag{1.18}$$

for some $p_1, p_2 \in (0,1]$. Now let $p \in [0,1]$ and $\mathsf{e} = p\mathsf{e}_1 + \bar{p}\mathsf{e}_2$. We have:

$$\begin{aligned} \mathsf{e} = p\mathsf{e}_1 + \bar{p}\mathsf{e}_2 &= p(p_1\mathsf{a} + \bar{p}_1\mathsf{b}_1) + \bar{p}(p_2\mathsf{a} + \bar{p}_2\mathsf{b}_2) \\ &= (pp_1 + \bar{p}p_2)\,\mathsf{a} + (p\bar{p}_1\mathsf{b}_1 + \bar{p}\bar{p}_2\mathsf{b}_2) \\ &= \lambda\mathsf{a} + \bar{\lambda}\left(\frac{p\bar{p}_1}{\bar{\lambda}}\mathsf{b}_1 + \frac{\bar{p}\bar{p}_2}{\bar{\lambda}}\mathsf{b}_2\right) = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b} \end{aligned} \tag{1.19}$$

where $\lambda = pp_1 + \bar{p}p_2$ and $\mathsf{b} = \frac{p\bar{p}_1}{\bar{\lambda}}\mathsf{b}_1 + \frac{\bar{p}\bar{p}_2}{\bar{\lambda}}\mathsf{b}_2$. Therefore, given any $\mathsf{a} \in \mathcal{E}$, we can find a $\mathsf{b} \in \mathcal{E}$ such that $\mathsf{e} = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b}$ for some $\lambda \in (0,1]$. This means the convex combination of internal points is an internal point, and I_ε is a convex set. □

It is an open question whether this result generalizes to infinite mixtures. That is, whether the infinite mixture of internal points is still an internal point. To generalize the previous proof to the infinite case, one would need to show that the infinite mixture of b_i converges. If conjecture 1.113 is true, that is if submixtures of convergent infinite mixtures converge, then the infinite mixture of b_i would converge as it is a submixture of e. We leave this as conjecture 1.114.

For topological vector spaces, the algebraic interior and the topological interior are related. For example, if A is a convex subset of a TVS with non-empty topological interior, then the algebraic and topological interior coincide. It is an open question how many of these results hold for topological convex spaces as well. As an example, we would like to prove the following.

TODO: reorganize

Conjecture 1.20. *Let $\mathcal{E}$ be an ensemble space. The set of boundary points $B_\mathcal{E}$ is not necessarily a closed set and therefore $I_\mathcal{E}$ is not necessarily an open set.*

Proof. Let $\mathcal{E} = \{p_i \in \ell^1 \mid \sum p_i = 1,\ p_i \in [0,1]\}$ be the set of probability distributions over countably infinitely many elements with the topology of ℓ^1. Let $\mathsf{e} \in I_\mathcal{E}$ be an interior point of $\mathcal{E}$. Since e is in the interior, it corresponds to a probability distribution p_i such that $p_i \in (0,1)$ for all i.

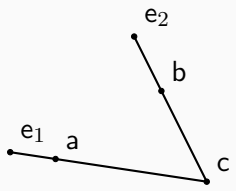
Let $B_r(\mathsf{e}) \subseteq \ell^1$ be an open ball of radius r centered around e. Since $\sum p_i = 1$, there must be a j such that $p_j < \frac{r}{2}$. Suppose, without loss of generality, that $j \neq 1$. Now consider the probability distribution given by $\lambda_i = (p_1 + p_j, p_2, \ldots, p_{j-1}, 0, p_{j+1}, \ldots)$. Since $\lambda_j = 0$, then λ_i is a boundary point of $\mathcal{E}$. The distance between the two distribution will be $\sum |p_i - \lambda_i| = |p_j| + |-p_j| = 2p_j < r$, which means $\lambda_i \in B_r(\mathsf{e})$. Therefore, the open ball around any internal point will contain a boundary point. Therefore the interior of $I_\mathcal{E}$ is the empty set, and $I_\mathcal{E}$ is not an open set. □

One may be able to show that the limit of a sequence of boundary points cannot be an interior point in the topological sense.

We now define the notions of separateness: two ensembles are separate, noted $\mathsf{a} \Perp \mathsf{b}$ if they do not have a common component. That is, they are not a mixture of a common ensemble. In classical ensemble spaces, this is equivalent to probability measures with disjoint support, which is the extension of the concept. Separateness is a useful concept to characterize the relationship between ensembles as it has some useful properties. Most of all, it is an irreflexive symmetric relation: nothing is separate from itself and if $\mathsf{a} \Perp \mathsf{b}$, then $\mathsf{b} \Perp \mathsf{a}$.[7]

[7] Note that orthogonality in inner product vector spaces is also an irreflexive symmetric relationship.

Definition 1.21. *Let $\mathcal{E}$ be an ensemble space and $\mathsf{a}, \mathsf{b} \in \mathcal{E}$. We say that they* ***have a common component*** *if we can find $\mathsf{c} \in \mathcal{E}$, the common component, such that $\mathsf{a} = p_1\mathsf{c} + \bar{p}_1\mathsf{e}_1$ and $\mathsf{b} = p_2\mathsf{c} + \bar{p}_2\mathsf{e}_2$ for some $\mathsf{e}_1, \mathsf{e}_2 \in \mathcal{E}$ and $p_1, p_2 \in (0, 1]$. Otherwise, we say they* ***have no common component****, or are* ***separate****, noted $\mathsf{a} \pitchfork \mathsf{b}$. Two ensembles have a common component in $A \subseteq \mathcal{E}$ if the common component can be found in A, and are separate in A if there is none. Two sets of ensembles $A, B \subseteq \mathcal{E}$ are separate if all the elements of one are separate from all the elements of the other. That is, $A \pitchfork B$ if $\mathsf{a} \pitchfork \mathsf{b}$ for all $\mathsf{a} \in A$ and $\mathsf{b} \in B$.*



Remark. If two ensembles a and b have a common component c, then the ensemble space contains a triangle where c is a vertex and a and b are points on the sides that connect to c.

Corollary 1.22. *The previous definitions obey the following:*

1. *every ensemble is a component of itself*
2. *if a is a component of b, then a and b have a common component and therefore they are not separate*
3. *separateness is an irreflexive symmetric relation*
4. *an ensemble is an extreme point if and only if it is separate from all other ensembles*
5. *an ensemble is an internal point if and only if it is not separate from any ensemble*
6. *if two ensembles are separate, then they are boundary points*

Proof. 1. Since by idempotence $\mathsf{e} = p\mathsf{e} + \bar{p}\mathsf{e}$ for any p, then every ensemble is a mixture of itself, and therefore it is a component of itself.

2. By idempotence, we can write $\mathsf{a} = p_1\mathsf{a} + \bar{p}_1\mathsf{a}$ for some $p_1 \in (0, 1]$. Since a is a component of b, we can write $\mathsf{b} = p_2\mathsf{a} + \bar{p}_2\mathsf{e}_2$ for some $p_2 \in (0, 1]$ and $\mathsf{e}_2 \in \mathcal{E}$. Therefore a and b have a as a common component.

3. Since every ensemble is a component of itself, every ensemble has a common component with itself and therefore is not separate from itself. This proves that separateness is irreflexive. The definition of common component is symmetric and therefore so is separateness.

4. An extreme point has only itself as a component, therefore it can have a common component only with itself.

5. Every ensemble is a component of an internal point, therefore every ensemble is not separate from an internal point

6. Since an internal point cannot be separate from any ensemble, then two ensembles that are separate are not internal points, and therefore are boundary points. □

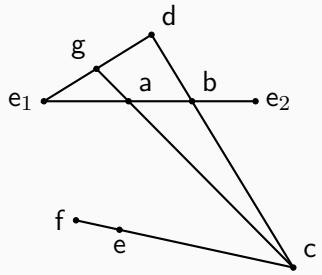
Proposition 1.23. *Let $\mathcal{E}$ be a discrete or continuous classical ensemble space and let $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ be two probability measures over the corresponding sample space X. Then $\mathsf{a} \pitchfork \mathsf{b}$ if and only if they have disjoint support.*

Proof. Note that the support of a convex combination of probability measures is the union of the support of the measures. Since we are restricting ourselves to probability measures that are absolutely continuous with respect to a measure μ, if a probability measure a has support U, any probability measure b with support on a compact subset $V \subseteq U$ such that $\mu(V) \neq 0$ is a component of a.

Let $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ be two probability measures over the sample space X that are absolutely continuous with respect to the corresponding μ. Suppose a and b have overlapping support. The we can find a compact subset U such that $\mu(U) \neq 0$ and is a subset of the intersection of their supports. Therefore, we can find a probability measure that is a component of both. Conversely, suppose they have disjoint support. Then they cannot have a common component, as the support of the common component would have to be a non-empty subset of both supports. □

A key property of separateness is its relationship with mixtures. If an ensemble is separate from a mixture of two elements, it is separate from both elements and all their mixtures.

Proposition 1.24 (Separateness extends to all mixtures)**.** *Let $\mathsf{e}, \mathsf{e}_1, \mathsf{e}_2 \in \mathcal{E}$. If e has no common component with a mixture of e_1 and e_2 then it has no common component with any mixture of e_1 and e_2 and with either e_1 or e_2. That is, if $\mathsf{e} \pitchfork p\mathsf{e}_1 + \bar{p}\mathsf{e}_2$ for some $p \in (0,1)$ then $\mathsf{e} \pitchfork p\mathsf{e}_1 + \bar{p}\mathsf{e}_2$ for all $p \in [0,1]$.*



Proof. Let $\mathsf{e} \pitchfork \mathsf{a} = p\mathsf{e}_1 + \bar{p}\mathsf{e}_2$ for some $p \in (0,1)$. Let $\mathsf{b} = \alpha\mathsf{e}_1 + \bar{\alpha}\mathsf{e}_2$ with $0 \leq \alpha < p$. As shown in the figure, suppose b is not separate from e. Then we can find $\mathsf{c} \in \mathcal{E}$ such that $\mathsf{b} = \beta\mathsf{c} + \bar{\beta}\mathsf{d}$ and $\mathsf{e} = \gamma\mathsf{c} + \bar{\gamma}\mathsf{f}$ for some $\mathsf{d}, \mathsf{f} \in \mathcal{E}$ and $\beta, \gamma \in (0,1)$.

Setting $\epsilon = \frac{p-\alpha}{\bar{\alpha}}$ and $\lambda = \bar{\epsilon}\beta$ we have:

$$
\begin{aligned}
\mathsf{a} &= p\mathsf{e}_1 + \bar{p}\mathsf{e}_2 = \left(p - \frac{\bar{p}}{\bar{\alpha}}\alpha\right)\mathsf{e}_1 + \frac{\bar{p}}{\bar{\alpha}}\alpha\mathsf{e}_1 + \frac{\bar{p}}{\bar{\alpha}}\bar{\alpha}\mathsf{e}_2 \\
&= \left(\frac{p\bar{\alpha} - \bar{p}\alpha}{\bar{\alpha}}\right)\mathsf{e}_1 + \frac{\bar{p}}{\bar{\alpha}}(\alpha\mathsf{e}_1 + \bar{\alpha}\mathsf{e}_2) = \left(\frac{p - p\alpha - \alpha + p\alpha}{\bar{\alpha}}\right)\mathsf{e}_1 + \frac{1 - p + \alpha - \alpha}{\bar{\alpha}}(\alpha\mathsf{e}_1 + \bar{\alpha}\mathsf{e}_2) \\
&= \frac{p-\alpha}{\bar{\alpha}}\mathsf{e}_1 + \left(1 - \frac{p-\alpha}{\bar{\alpha}}\right)(\alpha\mathsf{e}_1 + \bar{\alpha}\mathsf{e}_2) = \epsilon\mathsf{e}_1 + \bar{\epsilon}(\alpha\mathsf{e}_1 + \bar{\alpha}\mathsf{e}_2) = \epsilon\mathsf{e}_1 + \bar{\epsilon}\mathsf{b} \\
&= \epsilon\mathsf{e}_1 + \bar{\epsilon}(\beta\mathsf{c} + \bar{\beta}\mathsf{d}) = \bar{\epsilon}\beta\mathsf{c} + \epsilon\mathsf{e}_1 + \bar{\epsilon}\bar{\beta}\mathsf{d} = \lambda\mathsf{c} + \bar{\lambda}\mathsf{g}
\end{aligned}
$$

where $\mathsf{g} = \frac{1}{\lambda}\left(\epsilon\mathsf{e}_1 + \bar{\epsilon}\bar{\beta}\mathsf{d}\right)$. This means a and e have a common component, which is a contradiction. Therefore $\mathsf{e} \pitchfork \alpha\mathsf{e}_1 + \bar{\alpha}\mathsf{e}_2$ for all $\alpha \in [0, p]$.

We can repeat the argument switching e_1 with e_2 and find $\mathsf{e} \pitchfork \alpha\mathsf{e}_1 + \bar{\alpha}\mathsf{e}_2$ for all $\alpha \in [0, 1]$. □

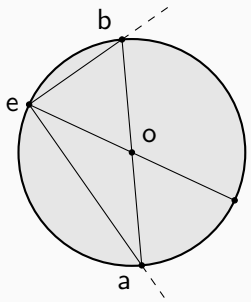
While separateness extends to mixtures, the converse property (i.e. mixtures preserve separateness) is not necessarily true. While it is true for classical spaces, it is not true for quantum spaces. This converse property is, in effect, a telltale of classicality.

Definition 1.25. *Let $\mathcal{E}$ be an ensemble space. We say that* ***mixtures preserve separateness in*** *$\mathcal{E}$ if $\mathsf{e} \pitchfork \mathsf{a}$ and $\mathsf{e} \pitchfork \mathsf{b}$ implies $\mathsf{e} \pitchfork p\mathsf{a} + \bar{p}\mathsf{b}$ for all $p \in [0, 1]$ and $\mathsf{e}, \mathsf{a}, \mathsf{b} \in \mathcal{E}$.*

Proposition 1.26. *Mixtures preserve separateness in discrete/continuous classical ensemble spaces.*

Proof. Let $\mathcal{E}$ be a discrete or continuous classical ensemble space. Let $\mathsf{e}, \mathsf{a}, \mathsf{b} \in \mathcal{E}$ such that $\mathsf{e} \pitchfork \mathsf{a}$ and $\mathsf{e} \pitchfork \mathsf{b}$. Then the support of e is disjoint from the support of both a and b. Since the support of a mixture of a and b is the union of the supports, e has disjoint support from every mixture of a and b. This means that mixtures preserve separateness in discrete/continuous classical ensemble spaces. □

Proposition 1.27. *Mixtures do not preserve separateness in quantum ensemble spaces*



Proof. Let $\mathcal{E}$ be a quantum ensemble space. As shown in the figure, let $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ be two orthogonal pure states and let $\mathsf{e} \in \mathcal{E}$ be another pure state that is the nontrivial superposition of the two. These three states will be extreme points of a Bloch ball. Since they are all pure states, they are all extreme points and therefore are pairwise separate. Consider $\mathsf{o} = \frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}$. This will be the center of the Bloch ball and will not be separate from e. Therefore mixtures do not preserve separateness in $\mathcal{E}$. □

Decomposability

Since mixture preserving separateness is a telltale of classicality, it is insightful to find an alternative characterization. One of the differences between classical and quantum ensemble spaces is that quantum ensembles allow multiple decomposition in terms of pure states. We see here that, in fact, the lack of multiple decomposition is equivalent to mixture preserving separateness.

Note that, since a continuous classical ensemble space has no extreme points, we have to

find a definition of multiple decomposability that makes no reference to the extreme points. Like for the definition of the entropy, we need to find a definition on mixtures of two elements that, when applied recursively, gives us the desired effect. The idea here is that we will always have multiple decompositions in terms of other ensembles, but in classical spaces we cannot have multiple decompositions where one element of the first decomposition has no common components with all elements of the second. So, if we start breaking an ensemble into separate components, while we may take different paths, we will reach the same final decomposition in terms of extreme points, if they exist.[8]

Definition 1.28. *An ensemble is* ***decomposable*** *if it can be expressed as a mixture of two distinct ensembles. An ensemble is* ***separately decomposable*** *if it can be expressed as a mixture of two separate ensembles. An ensemble is* ***multidecomposable*** *if it can be expressed as two decompositions where a component of one is separate from both components of the other. That is,* $\mathsf{e} = p\mathsf{a}_1 + \bar{p}\mathsf{a}_2 = \lambda\mathsf{b}_1 + \bar{\lambda}\mathsf{b}_2$ *and either* $\mathsf{a}_1 \perp\!\!\!\perp \mathsf{b}_j$ *or* $\mathsf{a}_2 \perp\!\!\!\perp \mathsf{b}_j$*. An ensemble is* ***monodecomposable*** *if it is not multidecomposable. An ensemble is* ***separately monodecomposable*** *if it is both separately decomposable and monodecomposable.*

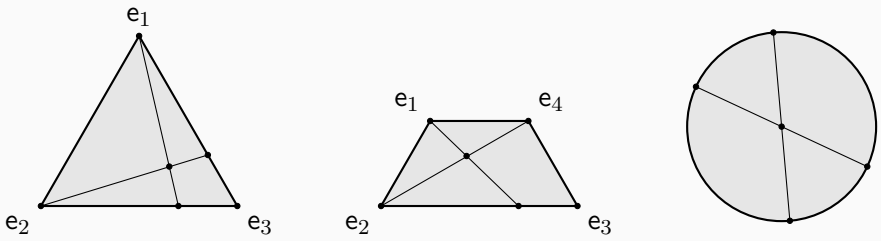


Figure 1.2: Examples for mono- and multidecomposability.

Remark. As shown in Fig. 1.2, take a classical discrete space for three points which is a triangle (simplex). The only three elements that are not decomposable are the extreme points. Mixtures of two points are decomposable and are also separately decomposable in only one way. Mixtures of three points are also separately decomposable, but in multiple ways: as a mixture of e_1 and a mixture of e_2 and e_3, or as a mixture of e_2 and a mixture of e_1 and e_3. Note, however, that they are not separately multidecomposable as the different components are not separate.

Take a Bloch ball for quantum mechanics. All the elements of the surface are not decomposable and they are all pairwise separate. The middle point can be seen as the equal mixture of any pair of opposite points. Therefore the middle point, as well as any other point not on the surface, is not only separately decomposable but also multidecomposable.

[8] There is an open issue as it is not clear whether we want to require actual separate decomposition or separate decomposition in the limit. In the continuous classical case, for example, it is not clear whether we should require probability densities to be continuous or not. If continuity is required, a probability density cannot be split into two probability densities with disjoint support without creating a discontinuity. However, we may think of a sequence of decomposition into three, one with support A, one with support B and one with support on both. In the limit, the mixing coefficient for the third becomes smaller and smaller, meaning that the first two approach a discontinuous distribution.

To see why we require only one component to be separate from the other two, consider the cut triangle. Here we can have multiple decompositions where not all elements are separate.

Corollary 1.29. *An ensemble* $\mathsf{e} \in \mathcal{E}$ *is an extreme point if and only if it not decomposable.*

Proof. If $\mathsf{e} \in \mathcal{E}$ is decomposable, then it has at least two distinct components, one of which must not be e. Since an extreme point has only itself as a component, then e is not an extreme point. Conversely, if e is an extreme point, it cannot be the mixture of two distinct components as e has only itself as a component. □

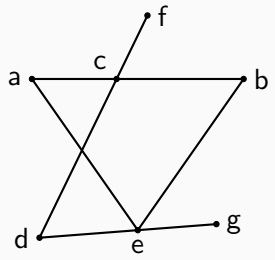
Definition 1.30. *An ensemble space is* ***separately decomposable*** *if every decomposable ensemble is separately decomposable,* ***multidecomposable*** *if every decomposable ensemble is multidecomposable,* ***monodecomposable*** *if every decomposable ensemble is monodecomposable and* ***separately monodecomposable*** *if it is both separately decomposable and monodecomposable.*

Proposition 1.31. *Discrete/continuous classical and quantum ensemble spaces are separately decomposable.*

Proof. For a classical ensemble space, only Dirac measures are not decomposable. For a discrete classical space, every measure that is not the Dirac measure is the mixture of two measures with disjoint support. For a continuous classical space, there are no extreme points, and any absolutely continuous measure is the mixture of two absolutely continuous measures with disjoint support. Since measures with disjoint support are separate, every decomposable ensemble is separately decomposable.

In quantum mechanics, every density operator is the mixture of its eigenstates. If the density operator does not correspond to a pure state, it will have at least two eigenstates. Every mixed state, then, is a mixture of two orthogonal ensembles. Since orthogonal ensembles cannot have a common component, every decomposable ensemble is separately decomposable. □

Proposition 1.32. *Given an ensemble space* $\mathcal{E}$*,* $\mathcal{E}$ *is monodecomposable if and only if mixtures preserve separateness in* $\mathcal{E}$*.*



Proof. Suppose mixtures do not preserve separateness in $\mathcal{E}$. Then, as shown in the figure, we can find $\mathsf{e}, \mathsf{a}, \mathsf{b}, \mathsf{c} \in \mathcal{E}$ such that $\mathsf{e} \perp \mathsf{a}$, $\mathsf{e} \perp \mathsf{b}$ and $\mathsf{e} \not\perp \mathsf{c} = p\mathsf{a} + \bar{p}\mathsf{b}$ for some $p \in (0, 1)$. Since $\mathsf{e} \not\perp \mathsf{c}$, we can find $\mathsf{d}, \mathsf{f}, \mathsf{g}$ such that $\mathsf{c} = \lambda\mathsf{d} + \bar{\lambda}\mathsf{f}$ and $\mathsf{e} = \mu\mathsf{d} + \bar{\mu}\mathsf{g}$. Since

separateness extends to all mixtures (1.24) and $\mathsf{a} \pitchfork \mathsf{e} = \mu\mathsf{d} + \bar{\mu}\mathsf{g}$, then $\mathsf{a} \pitchfork \mathsf{d}$. Similarly, $\mathsf{b} \pitchfork \mathsf{d}$, which means that c is separately multidecomposable. Therefore, if mixtures do not preserve separateness in $\mathcal{E}$, not all decomposable ensembles are monodecomposable which means $\mathcal{E}$ is not monodecomposable.

Now suppose mixtures do preserve separateness, and let $\mathsf{e} = p\mathsf{a}_1 + \bar{p}\mathsf{a}_2 = \lambda\mathsf{b}_1 + \bar{\lambda}\mathsf{b}_2$ be a decomposable ensemble. Since a_1 is a component of e, then $\mathsf{e} \pitchfork \mathsf{a}_1$. Since e is a mixture of b_1 and b_2 and mixtures preserve separateness, then either $\mathsf{a}_1 \pitchfork \mathsf{b}_1$ or $\mathsf{a}_1 \pitchfork \mathsf{b}_2$. Similarly, $\mathsf{a}_2 \pitchfork \mathsf{b}_1$ or $\mathsf{a}_2 \pitchfork \mathsf{b}_2$. Therefore e is not multidecomposable. Since this applies to all decomposable ensembles e, $\mathcal{E}$ is monodecomposable. □

These definitions may be enough to prove that every finite-dimensional separately monodecomposable convex space is a simplex. For the infinite case, it would be nice to compare this characterization to a Choquet simplex.

Conjecture 1.33. *A finite-dimensional (i.e. there is a set of finitely many elements whose hull has non-empty interior) separately monodecomposable convex space is a simplex.*

Convex subsets and convex hull

In many cases, we will need to discuss the sets that contain all their possible mixtures. One typically distinguishes two cases. A set is convex if it allows all possible finite mixtures. This may be too restrictive as it may not include all possible infinite mixtures. A set is closed and convex if it includes all finite mixtures and their topological limits. Given that infinite mixtures are the topological limit of finite mixtures, a closed convex set contains all infinite mixtures. However, not all topological limits can be expressed as infinite mixture. For example, on the real line 1 can be seen as a limit of points within the open interval $(0, 1)$, but not as infinite convex combination. Therefore we add the notion of σ-convex set, a set that is closed under infinite mixtures.[9]

Definition 1.34. *Let $\mathcal{E}$ be an ensemble space. We say $A \subseteq \mathcal{E}$ is **convex** if it closed under finite mixtures (i.e. $\mathsf{a}, \mathsf{b} \in A$ implies $p\mathsf{a} + \bar{p}\mathsf{b} \in A$ with $p \in [0, 1]$), **σ-convex** if it is closed under infinite mixtures (i.e. $\mathsf{a}_i \in A$ implies $\sum_i p_i \mathsf{a}_i \in A$ for all possible infinite mixtures) and **closed and convex** if it is both convex and topologically closed.*

Corollary 1.35. *A closed and convex set is σ-convex. A σ-convex set is convex.*

Given a set of ensembles A, we can ask for all ensembles that can be constructed from A. The hull of A is the set of all finite mixtures of A, the σ-hull of A is the set of all infinite mixtures of A and the closed hull of A is the set of all the topological limits of finite mixtures of A. Notably, the closed hull of A is equivalent to the topological closure of the hull of A.

Definition 1.36. *Let $A \subseteq \mathcal{E}$ be a subset of an ensemble space. The **convex hull** of A, noted* $\mathrm{hull}(A)$ *is the set of all finite mixtures of elements contained in A (i.e. it is the smallest convex set that contains A). The **σ-hull** of A, noted* $\mathrm{shull}(A)$ *is the set of all infinite mixtures of elements contained in A (i.e. it is the smallest σ-convex set that contains A).*

[9] The mathematical properties of σ-convex sets are yet to be explored.

*The **closed hull** of A, noted* $\text{chull}(A)$ *is the smallest closed convex set that contains A.*

Remark. Note that, given a set A, not all elements of $\text{chull}(A)$ can be understood as infinite mixtures. That is, we can have $\text{shull}(A) \subset \text{chull}(A)$. For example, let $\mathcal{E}$ be the line segment $[0,1]$ and consider the set $A = \left\{\frac{1}{2^i}\right\}_{i=0}^{\infty}$. Every point in $(0,1]$ can be expressed as a finite mixture of two elements of A, for example, 1 and any number smaller than the target number. However, zero cannot be expressed as a convex combination of positive numbers, and therefore it is not an infinite mixture of A. However, zero is the limit of the sequence, and therefore it will be in the topological closure of A. This shows that the difference between σ-hull and convex hull exists already in finite dimensions. The convex hull and the σ-hull, instead, are the same in finite dimensions because Carathodory's theorem allows us to rewrite any infinite convex combination into a finite one.

For an example in which all hulls are different, consider the space of probability distributions $\mathcal{E}$ over countably many elements $X = \{x_i\}_{i=1}^{\infty}$. Let $\mathsf{a}_{ip} = px_i + \bar{p}x_{i+1}$ and let $A = \{\mathsf{a}_{ip} \,|\, i \geq 1, p \in (0,1)\} \subset \mathcal{E}$ be the set of all non-trivial mixtures of pairs of consecutive elements. A probability distribution with support over the full X cannot be expressed as a finite convex combination of elements of A, and will therefore not be in the convex hull. However, it can be expressed as an infinite convex combination, and therefore it will be in the σ-hull. An element $x_i \in X$, is not in the σ-hull, but it will be in the closed hull, as $x_i = \lim_{p\to 1} px_i + \bar{p}x_{i+1} = \lim_{p\to 1} \mathsf{a}_{ip} \in \text{chull}(A)$.

Corollary 1.37. *Given $A \subseteq \mathcal{E}$,* $\text{hull}(A) \subseteq \text{shull}(A) \subseteq \text{chull}(A)$.

Proof. All finite mixtures are also infinite mixtures with $p_i = 0$ for all $i > n$ for some n. Therefore $\text{hull}(A) \subseteq \text{shull}(A)$. All infinite mixtures are topological limits of finite mixtures. Therefore $\text{shull}(A) \subseteq \text{chull}(A)$. □

Proposition 1.38. *All three hull operators are closures. That is,* hull *satisfies the following three properties:*

1. ***extensive**: $A \subseteq \text{hull}(A)$*
2. ***increasing**: $A \subseteq B \implies \text{hull}(A) \subseteq \text{hull}(B)$*
3. ***idempotent**: $\text{hull}(\text{hull}(A)) = \text{hull}(A)$*

and similarly do shull *and* chull.

Proof. 1. Every element of A is trivially a mixture of elements of A. Therefore $A \subseteq \text{hull}(A)$. Since $\text{hull}(A) \subseteq \text{shull}(A) \subseteq \text{chull}(A)$, $A \subseteq \text{shull}(A)$ and $A \subseteq \text{chull}(A)$ as well.

2. Let $\mathsf{e} \in \text{hull}(A)$. Then it is a finite mixture of some elements of A. Since $A \subseteq B$, then e is also the finite mixture of some elements of B and therefore $\mathsf{e} \in \text{hull}(B)$. The same logic applies to the σ-hull and closed hull replacing finite mixture with the appropriate operation.

3. Since $\text{hull}(\text{hull}(A))$ is the smallest convex subset that contains $\text{hull}(A)$, and since $\text{hull}(A)$ is a convex subset, then $\text{hull}(\text{hull}(A))$ must be $\text{hull}(A)$ since no smaller set can contain all elements of $\text{hull}(A)$. The same logic applies to the σ-hull and closed hull. □

Corollary 1.39. *A subset $A \subseteq \mathcal{E}$ is respectively convex/σ-convex/closed convex if and only if it is its own convex hull/σ-hull/closed hull.*

Proof. Let $A \subseteq \mathcal{E}$ be a convex subset. By 1.38 we have $A \subseteq \text{hull}(A)$. By definition of convex set, we have $\text{hull}(A) \subseteq A$. Therefore $A = \text{hull}(A)$. Conversely, let $A \subseteq \mathcal{E}$ be a set of ensembles not necessarily convex and let $A = \text{hull}(A)$. By definition, $\text{hull}(A)$ is closed under finite mixture and is therefore a convex subset. The same logic applies to σ-convex and closed convex sets with the respective hulls. □

Definition 1.40. *We note $\mathfrak{co}_{\mathcal{E}}$ the set of all convex subsets of $\mathcal{E}$, $\mathfrak{sco}_{\mathcal{E}}$ the set of all σ-convex subsets of $\mathcal{E}$ and $\mathfrak{cco}_{\mathcal{E}}$ the set of all closed convex subsets of $\mathcal{E}$.*

Proposition 1.41. *The sets $\mathfrak{co}_{\mathcal{E}}$, $\mathfrak{sco}_{\mathcal{E}}$ and $\mathfrak{cco}_{\mathcal{E}}$, as posets ordered by inclusion, are topped $\bigcap$-structures and therefore complete lattices.*

Proof. Theorem 7.3 in Davey and Priestley's "Introduction to Lattice and Order" states that, given a closure operator, the set of all closures, ordered by inclusion, is a topped $\bigcap$-structure and, therefore, a complete lattice. Since $\mathfrak{co}_{\mathcal{E}}$, $\mathfrak{sco}_{\mathcal{E}}$ and $\mathfrak{cco}_{\mathcal{E}}$ are closures, the theorem applies. □

Proposition 1.42. *The functions* hull, shull *and* chull *are continuous from above. That is, given a decreasing sequence $A_i \subseteq \mathcal{E}$, $\text{hull}(\lim_{i\to\infty} A_i) = \lim_{i\to\infty} \text{hull}(A_i)$. Similarly for* shull *and* chull.

Proof. The above proposition is a consequence of the fact that the hulls are closure operations and they generate an intersection structure. This means that the intersection of hulls is the hull of the intersections.

Let $A_i \subseteq \mathcal{E}$ be a decreasing sequence. That is, $A_{i+1} \subseteq A_i$. Then $A = \lim_{i\to\infty} A_i = \bigcap A_i$. Since hull is order preserving, $\text{hull}(A_i)$ is a decreasing sequence and $\lim_{i\to\infty} \text{hull}(A_i) = \bigcap \text{hull}(A_i)$. Moreover, $\text{hull}(A) \subseteq \text{hull}(A_i)$ for all i and therefore $\text{hull}(A) \subseteq \bigcap \text{hull}(A_i)$. Now let $\mathsf{e} \in \text{hull}(A)$. Then e is a convex combination of elements of A. Since every element of A is also an element of any A_i, then e is also a convex combination of elements of A_i for any i. Therefore $\mathsf{e} \in \text{hull}(A_i)$ for all i which means $\mathsf{e} \in \bigcap \text{hull}(A_i)$ and therefore $\text{hull}(A) \supseteq \bigcap \text{hull}(A_i)$. Thus we have that $\text{hull}(\lim_{i\to\infty} A_i) = \lim_{i\to\infty} \text{hull}(A_i)$.

Since we have only used closure properties of hull, the same reasoning applies to shull and chull since they are closures. □

Proposition 1.43. *The* hull *is continuous from below. That is, let $A_i \subseteq \mathcal{E}$ be an increasing sequence. Then* $\text{hull}(\lim_{i\to\infty} A_i) = \lim_{i\to\infty} \text{hull}(A_i)$.

Proof. Let $A_i \subseteq \mathcal{E}$ be an increasing sequence. That is, $A_{i+1} \supseteq A_i$. Then $A = \lim_{i\to\infty} A_i = \bigcup A_i$. Since hull is an increasing function, $\text{hull}(A) \supseteq \text{hull}(A_i)$ for all i and therefore $\text{hull}(A) \supseteq \bigcup \text{hull}(A_i)$. Now let $\mathsf{e} \in \text{hull}(A)$. Then e is a convex combination of finitely many elements a_j of A. Since A is the union of all A_i, each a_j will be in some A_i. Since the sequence of A_i is increasing, and there are only finitely many a_j, we will find an i such that $\mathsf{a}_j \in A_i$ for all j. This means that e is a convex combination of elements of A_i and therefore $\mathsf{e} \in \text{hull}(A_i) \subseteq \text{hull}(A)$. Therefore $\text{hull}(A) = \bigcup \text{hull}(A_i)$ which means $\text{hull}(\lim_{i\to\infty} A_i) = \lim_{i\to\infty} \text{hull}(A_i)$. □

Remark. Note that shull and chull are not, in general, continuous from below. This is because, in general, the union of closures is not the closure of the union. This is true, in particular, with topological closures, which is part of the definition of chull.

For example, consider the sequence $A_i = \left[0, 1 - \frac{1}{i}\right) \subseteq \mathbb{R}$. These are convex sets therefore their closed hull is simply their topological closure. That is, $\text{chull}(A_i) = \left[0, 1 - \frac{1}{i}\right]$. We have $\lim_{i\to\infty} A_i = \bigcup A_i = [0, 1)$ and $\lim_{i\to\infty} \text{chull}(A_i) = \bigcup \text{chull}(A_i) = [0, 1)$ which is not a closed set and therefore different from $\text{chull}(A) = [0, 1]$. The closed hull of the limit is not the limit of the convex hull, even in a finite-dimensional space.

For the σ-hull, we need an infinite-dimensional example. Conceptually, we are using the fact that a uniform distribution over the whole $[0, 1]$ is the infinite convex combination of uniform distributions over countably many sets that cover the whole $[0, 1]$. Let $\mathcal{E}$ be the space of probability measures defined over $[0, 1] \subseteq \mathbb{R}$. Let $\{\mathsf{a}_i\}_{i=1}^{\infty}$ be the sequence of uniform distributions over $\left[\frac{1}{i+1}, \frac{1}{i}\right]$. Let e be the uniform distribution over $[0, 1]$. We have $\mathsf{e} = \sum \frac{1}{i(i+1)} \mathsf{a}_i$ where $\sum \frac{1}{i(i+1)} = 1$. Therefore e is the countable convex combination of a_i. Let $A_j = \{\mathsf{a}_i \,|\, i \leq j\}$. Note that $\mathsf{a}_i \Perp \mathsf{a}_j$ for all $i \neq j$. Therefore, $\{\mathsf{a}_i\}$ are exactly all extreme points of $\text{shull}(\bigcup A_j)$. This means that $\mathsf{e} \notin \text{shull}(A_j)$ for all j while $\mathsf{e} \in \text{shull}(\bigcup A_j)$. Therefore the σ-hull of the limit is not the limit of the σ-hulls.

Proposition 1.44. *The topological closure of the hull is a convex set and, therefore, the closed hull.*

Proof. Let A be a convex set and $\bar{A}$ its topological closure. Let $\mathsf{a}_i, \mathsf{b}_i \in A$ be two sequences that converge to $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ respectively. We have $\mathsf{a}, \mathsf{b} \in \bar{A}$ since they are topological limits. Consider $\mathsf{e}_i = p\mathsf{a}_i + \bar{p}\mathsf{b}_i$. Since mixing is continuous, we have:

$$\mathsf{e} = p\mathsf{a} + \bar{p}\mathsf{b} = p \lim_{i\to\infty} \mathsf{a}_i + \bar{p} \lim_{i\to\infty} \mathsf{b}_i = \lim_{i\to\infty} (p\mathsf{a}_i + \bar{p}\mathsf{b}_i) = \lim_{i\to\infty} \mathsf{e}_i. \tag{1.45}$$

But e_i are finite mixtures of elements of A, and therefore $\mathsf{e}_i \in A$ is a sequence of elements of A. The sequence converges, e is the limit of a sequence of elements of A and therefore $\mathsf{e} \in \bar{A}$. That is, the topological closure of a convex set is also a convex set. But this means that $\bar{A}$ is a closed convex set. Since any closed convex set that contains A will also need to contain $\bar{A}$, $\bar{A}$ is the closed hull of A.

Let $A \subset \mathcal{E}$ be a subset not necessarily convex. Then $\text{hull}(A)$ will be convex. Therefore its topological closure will be the closed hull of A. □

There may be a relationship between the algebraic notion of σ-convexity and the topological notion of interior. For example, let U be a convex open set. While it clearly cannot be closed convex, is it σ-convex? The idea is that convexity can only return points that are "inside" the set, and σ-convexity is required to fill in all the limits. If that is true, it would be natural to look at some sort of converse. Clearly, not all σ-convex sets are open, since all closed convex sets are also σ-convex. The question becomes whether, for σ-convex sets, the notion of algebraic boundary and topological boundary coincides. These questions raise the following conjectures.

Conjecture 1.46. *Let $U \subseteq \mathcal{E}$ be a convex open set. Then U is σ-convex.*

Definition 1.47. *Given a set $A \subset \mathcal{E}$, $\mathsf{a} \in A$ is an internal point of A if for any $\mathsf{e} \in \mathcal{E}$ we can*

find $\mathsf{b} \in A$ *such that* $\mathsf{a} = p\mathsf{e} + \bar{p}\mathsf{b}$ *for some* $p \in (0, 1]$.

Conjecture 1.48. *Let* $A \subseteq \mathcal{E}$ *and let* $\mathsf{a} \in \text{chull}(A)$ *be an internal point of* $\text{chull}(A)$. *Then* $\mathsf{a} \in \text{shull}(A)$.

Conjecture 1.49. *Let* $A \subseteq \mathcal{E}$ *be a* σ*-convex set. Then* $\mathsf{a} \in A$ *is an internal point of* A *if and only if it is an interior point of* A.

Remark. Note that if A is not convex, this is clearly not true. For example, let $A \subseteq \mathbb{R}^2$ be an annulus (i.e. the region of two concentric circles). The points on the inner circle are internal points of A according to the definition, but they are not interior points. An internal point of A for a σ-convex set is guaranteed to be surrounded by an open interval along any direction. The question, as usual, is if this is enough to fit an open set.

Convex supremum

Later in the chapter, we will need to create the non-additive generalization for probability measures and state counting measures. It turns out that both constructions can be understood as instances of a more general construction that starts with a function $f(\mathsf{e})$ of elements of the ensemble space and generates a function $cs_f(A)$ of sets of ensembles by asking what is the highest value of f that is reachable by a mixture of elements of A. We are going to show that this construction alone presents many nice mathematical properties.

Definition 1.50. *Given a function* $f : \mathcal{E} \to \mathbb{R}$, *a* ***convex supremum*** *of* f *is a set function* $cs_f : 2^{\mathcal{E}} \to [-\infty, +\infty]$ *such that* $cs_f(A) = \sup(f(\text{hull}(A)) \cup \{f_\varnothing\})$, *with* $f_\varnothing \leq \inf(f(\mathcal{E}))$, *that returns the highest value of* f *reachable by convex combinations of* A.

Proposition 1.51. *For any* f, *the convex supremum* cs_f *has the following properties*

1. *range of* f: $cs_f(A) \in [f_\varnothing, \sup(f(\mathcal{E}))]$
2. *increasing:* $A \subseteq B \implies cs_f(A) \leq cs_f(B)$
3. *continuous from below: for any increasing sequence* $A_i \subseteq \mathcal{E}$, $cs_f(\lim_{i\to\infty} A_i) = \lim_{i\to\infty} cs_f(A_i)$.

Proof. 1. If A is empty, $cs_f(A) = f_\varnothing$. If A is non-empty, $cs_f(A)$ is the supremum of the subset of values returned by $f(\text{hull}(A))$, which means $f_\varnothing \leq \inf(f(\mathcal{E})) \leq cs_f(A) \leq \sup(f(\mathcal{E}))$. Therefore $cs_f(A) \in [f_\varnothing, \sup(f(\mathcal{E}))]$ for all A.

2. If $A = B = \varnothing$, then $cs_f(A) = cs_f(B)$. If $B \neq \varnothing$, $cs_f(A) = f_\varnothing \leq \inf(\mathcal{E}) \leq cs_f(B)$. In the last case, since the hull is an increasing function, the image of a set through a map is an increasing function and the supremum is an increasing function, the convex supremum is an increasing function.

3. Let $A_i \subseteq \mathcal{E}$ be an increasing sequence and $A = \bigcup A_i$. Since cs_f is increasing, $cs_f(A_i)$ is also an increasing sequence. Since hull is continuous from below, we have:

$$\begin{aligned} cs_f(A) &= \sup(f(\text{hull}(A)) \cup \{f_\varnothing\}) = \sup(f(\bigcup \text{hull}(A_i)) \cup \{f_\varnothing\}) \\ &= \sup(\bigcup f(\text{hull}(A_i)) \cup \{f_\varnothing\}) = \sup(\bigcup(\sup(f(\text{hull}(A_i)) \cup \{f_\varnothing\}))) \\ &= \sup(\bigcup cs_f(A_i)). \end{aligned} \tag{1.52}$$

But since $cs_f(A_i)$ is increasing, the supremum is exactly the limit. Therefore the convex supremum is continuous from below. □

Remark. Note that the convex supremum is not continuous from above. Let $\mathcal{E}$ be a disc embedded in $\mathbb{R}^2$. Let $f : \mathcal{E} \to \mathbb{R}$ be a non-trivial linear function. Then f will have a maximum and a minimum on two opposite points a and b on the circle that encloses the disc. Take a line that divides the disc in two halves, leaving the minimum and the maximum on different halves. Then over that line f will take the minimum value $m > f(\mathsf{b})$ on one of the extreme points of the line. Consider a countable collection $\{\mathsf{e}_i\}$ of points over that line and let $A_j = \{\mathsf{e}_i \,|\, i \geq j\}$. This is a decreasing sequence of infinite sets, where we are taking one point out at a time. We have $f(\mathsf{e}_i) \geq m$ for all i and therefore $cs_f(A_j) \geq m$ for all j, which means $\lim_{j\to\infty} cs_f(A_j) \geq m$. However, $\bigcap A_j = \varnothing$ and $cs_f(\varnothing) = f_\varnothing \leq \inf(f(\mathcal{E})) = f(\mathsf{b}) < m$. This means that, in general, $cs_f(\lim_{j\to\infty} A_j) \neq \lim_{j\to\infty} cs_f(A_j)$ for a decreasing sequence.

Proposition 1.53. *Let $A \subseteq \mathcal{E}$ and $f : \mathcal{E} \to \mathbb{R}$ be a continuous function, then $cs_f(A) = cs_f(\text{shull}(A)) = cs_f(\text{chull}(A))$.*

Proof. The proposition is true if $A = \varnothing$, since $\varnothing = \text{hull}(\varnothing) = \text{shull}(\varnothing) = \text{chull}(\varnothing)$.

Now let $A \neq \varnothing$. Note that $\text{chull}(A)$ is a convex set, meaning $\text{hull}(\text{chull}(A)) = \text{chull}(A)$. This means that we are looking for the difference between the supremum of f over the hull and the closed hull. Suppose $\mathsf{e} \in \text{chull}(A)$ but $\mathsf{e} \notin \text{hull}(A)$. Still, e will be the limit of a sequence of $\mathsf{e}_i \in \text{hull}(A)$. Since f is continuous, then $f(\mathsf{e})$ is the limit of the sequence $f(\mathsf{e}_i)$, for which $f(\mathsf{e}_i) \leq \sup(f(\text{hull}(A)))$ for all i. Therefore, $f(\mathsf{e}) \leq \sup(f(\text{hull}(A))) = cs_f(A)$. Since $f(\mathsf{e}) \leq cs_f(A)$ for all $\mathsf{e} \in \text{chull}(A)$, then $\sup(f(\text{chull}(A))) = cs_f(\text{chull}(A)) \leq cs_f(A)$. Since cs_f is increasing and $\text{hull}(A) \subseteq \text{chull}(A)$, $cs_f(A) \leq cs_f(\text{chull}(A))$. Therefore $cs_f(A) = cs_f(\text{chull}(A))$.

Note that $A \subseteq \text{shull}(A) \subseteq \text{chull}(A)$ and cs_f is an increasing function. Therefore $cs_f(A) \leq cs_f(\text{shull}(A)) \leq cs_f(\text{chull}(A)) = cs_f(A)$ which means $cs_f(A) = cs_f(\text{shull}(A))$. □

Note that a measure is a monotonic set function continuous from below that is also non-negative and additive. It is easy to show that the convex supremum is non-negative if and only if f is non-negative. Additivity, instead, is more complicated as it needs to be recovered on the lattice of subspaces.

Corollary 1.54. *A convex supremum of f is non-negative if and only if f is non-negative and $f_\varnothing \geq 0$.*

Proof. Since the range of cs_f is $[f_\varnothing, \sup(f(\mathcal{E}))]$, cs_f is non-negative if and only if $f_\varnothing \geq 0$. Since $f_\varnothing \leq \inf(f(\mathcal{E}))$, $f_\varnothing \geq 0$ only if f is non-negative. □

1.4 Axiom of entropy

In this section we will see how characterizing the variability of the elements within an ensemble leads to the standard notion of entropy. The entropy will have to satisfy few basic properties justified by the notion of variability which will be enough to recover the usual Shannon formula. It will also provide us with another basic notion to compare ensembles: two ensembles are mutually exclusive, or orthogonal, if they have no elements in common, which means the

entropy will maximally increase during mixing. The interaction between orthogonality and separateness is enough to understand the differences between classical and quantum systems.

We saw that our basic notion of state is that of an ensemble. Since an ensemble represents all possible preparations of equivalent systems prepared according to the same procedure, we want to characterize how much those instances are different from each other. To be truly general, we want to be able to make this characterization without assuming what the individual instances are. First of all, the variability should be an experimentally well-defined quantity, and it should therefore be a continuous function in the natural topology. It should also be compatible with statistical mixing. Intuitively, the variability cannot decrease during mixing which makes the entropy a strictly concave function. Variability will have a maximal increase when mixing ensembles that are mutually exclusive: no instance of one can be confused with an instance of the other. In that case, the increase of the variability is fully characterized by the mixing coefficients. We say that two mutually exclusive ensembles are orthogonal, as this property will correspond mathematically to orthogonality in the space of distributions. Lastly, if one ensemble has no instances in common with other two, then it has no instance in common with any mixture of the two.

Thinking of ensembles and the variability of their instances, then, makes it clear what entropy is and why it has these properties. We are not concerned, at this point, what the source of the variability is. We just know that it is something that we need to characterize.

Axiom 1.55 (Axiom of entropy)**.** *Every ensemble is associated with an* ***entropy*** *which quantifies the variability of the instances within an ensemble. Formally, an ensemble space $\mathcal{E}$ is equipped with a function $S : \mathcal{E} \to \mathbb{R}$, defined up to a positive multiplicative constant representing the unit numerical value. The entropy has the following properties:*

- ***Continuity***[a]
- ***Strict concavity****: $S(p\mathsf{a} + \bar{p}\mathsf{b}) \geq pS(\mathsf{a}) + \bar{p}S(\mathsf{b})$ with the equality holding if and only if $\mathsf{a} = \mathsf{b}$*
- ***Upper variability bound****: there exists a universal function $I(p, \bar{p})$ (i.e. the same for all ensemble spaces) such that $S(p\mathsf{a} + \bar{p}\mathsf{b}) \leq I(p, \bar{p}) + pS(\mathsf{a}) + \bar{p}S(\mathsf{b})$; if the equality holds, a and b are* ***mutually exclusive*** *or* ***orthogonal****, noted $\mathsf{a} \perp \mathsf{b}$*
- ***Mixtures preserve orthogonality****:*[b] *$\mathsf{a} \perp \mathsf{b}$ and $\mathsf{a} \perp \mathsf{c}$ if and only if $\mathsf{a} \perp p\mathsf{b} + \bar{p}\mathsf{c}$ for any $p \in (0, 1)$*

Justification. The entropy quantifies the variability of the instances within an ensemble. Since the ensemble represents a collection of preparations of equivalent systems, and since each instance will in general be potentially different, it is legitimate to ask how much variability there is among the different instances. We are assuming that the entropy is a quantity (i.e. a linearly ordered property), meaning that it is always meaningful to tell whether one ensemble has more variability than another. If this is the case, the later requirements of continuity and strict concavity will force the entropy to be a real-valued quantity. This is because the variability will change under statistical mixtures, and since statistical mixtures are performed with real-valued coefficients, the variability will have to be a real-valued quantity. While we strongly suspect that it is not conceptually possible to have a characterization of variability that is not linearly ordered, as this would entail

ranges of the same variable that can be potentially mapped to each other without a clear mapping of their variability, we do not yet have a tight argument. Therefore we are not able to fully justify entropy's linear ordering at this time, and the linear ordering of the entropy should be considered an assumption. Provided that assumption, we are justified to assume the existence of a real-valued function that returns the entropy, a measure of variability of the ensemble.

Since the entropy is a real-valued quantity, it will have a corresponding unit. This unit is independent from all other units, and therefore the overall structure of the ensemble space must be independent of this choice. Mathematically, the physical dimension of the unit is not captured, just its numeric value. A change of unit may change the numeric value by a multiplicative constant. Since variability is an ordered quantity, we want the change of units to respect the ordering and therefore it should be a positive multiplicative constant. This justifies that the entropy function is defined up to a positive multiplicative constant.

Note that the additivity of the entropy over independent systems fixes the zero. If $S_{AB} = S_A + S_B$, in fact, one can't rescale all three terms by an additive factor and preserve the relationship.

The variability, in the end, will have physical consequences, and it will therefore be measurable, thus experimentally verifiable: it will have to be a topologically continuous function. Moreover, small changes in the ensemble should produce small changes in the variability, which justifies analytical continuity. We are therefore justified to assume continuity of the entropy.

Suppose we have two ensembles and we perform a statistical mixture. There are going to be three sources of variability: the two ensembles and the random choice between them. The total contribution from the original ensembles will be the average variability of the original ensembles. This is increased by the variability introduced by the random choice, which is always a positive contribution. Therefore the final variability cannot be less than the average of the original ensembles. That is, $S(p\mathsf{a} + \bar{p}\mathsf{b}) \geq pS(\mathsf{a}) + \bar{p}S(\mathsf{b})$. If we are mixing an ensemble with itself, this is equivalent to just choosing from the original ensemble, therefore the variability will not increase. Conversely, if the variability stays the same, it means that the random choice does not increase the variability, and therefore we must be choosing between equivalent ensembles. Therefore we are justified to assume that entropy is strictly concave.

On the other hand, the variability cannot increase arbitrarily during mixture. The maximum variability will be given when the two ensembles are mutually exclusive, when an instance of the first ensemble cannot be produced by the second ensemble. That is, a single instance is enough to determine whether we have the first ensemble or the second. In this case, the variability is increased by the variability of the random choice, which must depend only on the mixture coefficient, and not the nature of the ensembles themselves. That is, $S(p\mathsf{a}+\bar{p}\mathsf{b}) \leq I(p,\bar{p})+pS(\mathsf{a})+\bar{p}S(\mathsf{b})$ is the upper variability bound, which is saturated if and only if a and b are mutually exclusive. The actual function I is left unspecified and, as we show in proposition 1.59, it will correspond to the Shannon entropy as it is the only indicator of variability that will satisfy the axiom of entropy. This justifies the upper variability bound.

Now suppose ensembles a and b are mutually exclusive and so are a and c. That is,

an instance of a cannot ever be produced by either b or c. Then an instance of a cannot be produced by a mixture of b and c, since ultimately a mixture of b and c will return an instance of one of the two. Therefore a and any mixture of b and c are mutually exclusive. The argument works in reverse as well: if an instance of a cannot be produced by a mixture of b and c, then it cannot be produced by either. This justifies mixtures preserve orthogonality. □

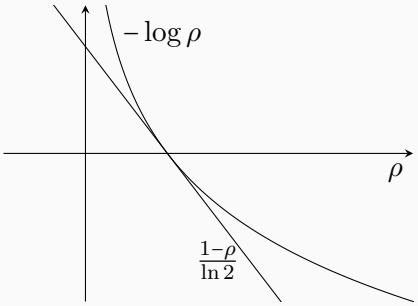
[a] Currently, we are imposing that the entropy is continuous. There may be a chance that this requirement is redundant, as strict concavity and the upper variability bound may already impose this. We have found proofs that show that real-valued convex/concave functions of real values are continuous. These proofs fail at the extreme points, but the upper variability bound may fix this. Another open question is whether differentiability is also an independent requirement.

[b] It is unclear whether "mixtures preserve orthogonality" is an independent axiom. Intuitively, the following argument tells us that it is. Take the 2-dimensional simplex (i.e. a triangle) that represents a classical discrete probability space over three elements. Take the standard entropy, which will satisfy "mixtures preserve orthogonality". This is because the middle point has entropy $\log 3$. We can imagine redefining the entropy so that it is a little bit lower in the center but it is unchanged on the sides. However, one needs to provide an actual example and show that it satisfies all axioms. It may also be that only one direction of the implication is an independent axiom. That is, that orthogonality with the components implies orthogonality with the mixtures. This is what does not hold for separateness.

As with the other axioms, we should now verify that the axiom of entropy is satisfied by the standard cases.

Proposition 1.56. *Discrete classical ensemble spaces, continuous classical ensemble spaces and quantum ensemble spaces satisfy the axiom of entropy.*

Proof. Let's first look at the classical continuous case. Every ensemble is represented by a distribution $\rho(x)$ with $\int_X \rho(x)d\mu = 1$. The entropy is given by $S(\rho) = -\int_X \rho \log \rho d\mu$. This is a continuous function of ρ.

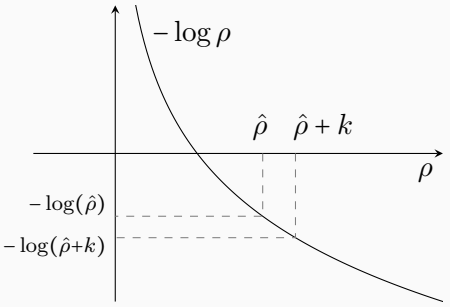


Recall that throughout this chapter, all logarithms are base two. To show strict concavity, note that, as shown in the figure, $-\log\rho \geq \frac{1-\rho}{\ln 2}$ with the equality holding if and only

if $\rho = 1$. We have

$$\begin{aligned}
S(p\rho_1 + \bar{p}\rho_2) &= -\int_X (p\rho_1 + \bar{p}\rho_2) \log (p\rho_1 + \bar{p}\rho_2)\, d\mu \\
&= -\int_X p\rho_1 \log (p\rho_1 + \bar{p}\rho_2)\, d\mu - \int_X \bar{p}\rho_2 \log (p\rho_1 + \bar{p}\rho_2)\, d\mu \\
&= -\int_X p\rho_1 \log \frac{p\rho_1 + \bar{p}\rho_2}{\rho_1} d\mu - \int_X p\rho_1 \log \rho_1 d\mu \\
&- \int_X \bar{p}\rho_2 \log \frac{p\rho_1 + \bar{p}\rho_2}{\rho_2} d\mu - \int_X \bar{p}\rho_2 \log \rho_2 d\mu \\
&\geq \int_X p\rho_1 \frac{1}{\ln 2} \left(1 - \frac{p\rho_1 + \bar{p}\rho_2}{\rho_1}\right) d\mu - p \int_X \rho_1 \log \rho_1 d\mu \\
&+ \int_X \bar{p}\rho_2 \frac{1}{\ln 2} \left(1 - \frac{p\rho_1 + \bar{p}\rho_2}{\rho_2}\right) d\mu - \bar{p} \int_X \rho_2 \log \rho_2 d\mu \\
&= \frac{p}{\ln 2} \left[\int_X \rho_1 d\mu - \int_X (p\rho_1 + \bar{p}\rho_2)\, d\mu\right] + pS(\rho_1) \\
&+ \frac{\bar{p}}{\ln 2} \left[\int_X \rho_2 d\mu - \int_X (p\rho_1 + \bar{p}\rho_2)\, d\mu\right] + \bar{p}S(\rho_2) \\
&= \frac{p}{\ln 2} [1 - 1] + pS(\rho_1) + \frac{\bar{p}}{\ln 2} [1 - 1] + \bar{p}S(\rho_2) \\
&= pS(\rho_1) + \bar{p}S(\rho_2)
\end{aligned} \tag{1.57}$$

The equality holds if and only if $\frac{p\rho_1 + \bar{p}\rho_2}{\rho_1} = 1$ which is exactly when $\rho_1 = \rho_2$.



For the upper bound, as shown in the figure, note that the logarithm is a strictly increasing function, and therefore $-\log(\hat{\rho} + k) \leq -\log(\hat{\rho})$ for any $k \geq 0$, with equality

holding if and only if $k = 0$. We have

$$\begin{aligned} S(p\rho_1 + \bar{p}\rho_2) &= -\int_X (p\rho_1 + \bar{p}\rho_2) \log (p\rho_1 + \bar{p}\rho_2)\, d\mu \\ &= -\int_X p\rho_1 \log (p\rho_1 + \bar{p}\rho_2)\, d\mu - \int_X \bar{p}\rho_2 \log (p\rho_1 + \bar{p}\rho_2)\, d\mu \\ &\leq -\int_X p\rho_1 \log p\rho_1 d\mu - \int_X \bar{p}\rho_2 \log \bar{p}\rho_2 d\mu \\ &= -\int_X p\rho_1 \log p d\mu - \int_X p\rho_1 \log \rho_1 d\mu - \int_X \bar{p}\rho_2 \log \bar{p} d\mu - \int_X \bar{p}\rho_2 \log \rho_2 d\mu \\ &= -p \log p \int_X \rho_1 d\mu - p \int_X \rho_1 \log \rho_1 d\mu - \bar{p} \log \bar{p} \int_X \rho_2 d\mu - \bar{p} \int_X \rho_2 \log \rho_2 d\mu \\ &= -p \log p - \bar{p} \log \bar{p} + pS(\rho_1) + \bar{p} S(\rho_2) \end{aligned} \tag{1.58}$$

The equality holds if and only if $\rho_2 = 0$ wherever $\rho_1 \neq 0$ and $\rho_1 = 0$ wherever $\rho_2 \neq 0$. That is, the equality holds if and only if the two distributions have disjoint support. Therefore orthogonal distributions are exactly distributions with disjoint support.

Suppose ρ_4 has disjoint support from $\rho_1 = p\rho_2 + \bar{p}\rho_3$, then it has disjoint support from ρ_2 and ρ_3 because the support of ρ_1 is the union of the supports of ρ_2 and ρ_3. Conversely, if ρ_4 has disjoint support from both ρ_2 and ρ_3, then ρ_4 has disjoint support from ρ_1 as well. Therefore mixtures preserve orthogonality.

All these arguments are valid for discrete classical ensemble spaces, changing integrals to sums.

For the quantum case, we haven't found a short proof that does not require defining the KL divergence and the entropy for a joint distribution. The result, however, is generally known and can be found, for example, in Nielsen and Chuang. □

Uniqueness of entropy of mixing coefficients

The axiom of entropy imposes only the existence of a universal function $I(p, \bar{p})$ without specifying its functional form. We are now going to show that the functional form is actually already fixed by the axiom: the Shannon entropy $-p \log p - \bar{p} \log \bar{p}$ is in fact the only function that is satisfies the axiom. The result is achieved by calculating the entropy of the same final ensemble as decomposed into a mixture in different ways. Since the final entropy should not care about how the mixture is performed, $I(p, \bar{p})$ must satisfy some properties that lead to the final results.

We note that the Shannon entropy in our framework does not represent an absolute entropy, but the maximal entropy increase during mixing. This is why we are able to keep the framework general. In fact, note that the proof does not technically know whether we are in a classical or quantum ensemble space, or even in an ensemble space of a new possible theory.

Theorem 1.59 (Uniqueness of entropy)**.** *If there exists an ensemble space with an infinite set of orthogonal ensembles, then the entropy of the coefficients $I(p, \bar{p})$ is the Shannon entropy. That is, $I(p, \bar{p}) = -\kappa (p \log p + \bar{p} \log \bar{p})$ where $\kappa > 0$ is the arbitrary multiplicative constant for the entropy. For a mixture of arbitrarily many elements, $I(\{p_i\}) = -\kappa \sum_i p_i \log p_i$.*

Proof. Since the upper entropy bound has to be the same for all spaces, let us assume that $\mathcal{E}$ is such that it contains countably many orthogonal ensembles $\{\mathsf{e}_l\}_{l=1}^{\infty}$. Since mixtures preserve orthogonality, for any convex combinations $\sum_i p_i \mathsf{a}_i$ of finitely many $\{\mathsf{a}_i\}_{i=1}^{n} \subset \{\mathsf{e}_l\}_{l=1}^{\infty}$, we have $S(\sum_i p_i \mathsf{a}_i) = I_n(p_1, p_2, \ldots, p_n) + \sum_i p_i S(\mathsf{a}_i)$ where $I_n : [0,1]^n \to \mathbb{R}$ is a function of the coefficients only. Note that, given commutativity, the order of the p_i does not matter and, since the coefficients can be zero, we must have $I_n(p_1, p_2, \ldots, p_n) = I_{n+1}(p_1, p_2, \ldots, p_n, 0)$. Therefore we can think of I as a function of the coefficients and write $I(\{p_i\}_{i=1}^{n})$. Note that since we can write $I(\{p_i\}_{i=1}^{n}) = S(\sum_i p_i \mathsf{a}_i) - \sum_i p_i S(\mathsf{a}_i)$, and both entropy and mixing are continuous, then I is continuous.

We now show that $I\left(\left\{\frac{1}{n}\right\}_{i=1}^{n}\right) = \kappa \log n$ with $\kappa > 0$. That is, the maximum increase of entropy for a uniform distribution is proportional to the logarithm of the number of cases. Pick two positive integers $n, m \in \mathbb{Z}^+$. Pick nm elements $\mathsf{a}_{jk} \in \{e_l\}_{l=1}^{\infty}$ where $1 \le j \le n$ and $1 \le k \le m$. We have:

$$\begin{aligned} S\left(\sum_{j=1}^{n}\sum_{k=1}^{m}\frac{1}{n}\frac{1}{m}\mathsf{a}_{jk}\right) &= I\left(\left\{\frac{1}{n}\frac{1}{m}\right\}_{i=1}^{nm}\right) + \sum_{j=1}^{n}\sum_{k=1}^{m}\frac{1}{n}\frac{1}{m}S\left(\mathsf{a}_{jk}\right) \\ = S\left(\sum_{j=1}^{n}\frac{1}{n}\sum_{k=1}^{m}\frac{1}{m}\mathsf{a}_{jk}\right) &= I\left(\left\{\frac{1}{n}\right\}_{i=1}^{n}\right) + \sum_{j=1}^{n}\frac{1}{n}S\left(\sum_{k=1}^{m}\frac{1}{m}\mathsf{a}_{jk}\right) \\ &= I\left(\left\{\frac{1}{n}\right\}_{i=1}^{n}\right) + \sum_{j=1}^{n}\frac{1}{n}\left(I\left(\left\{\frac{1}{m}\right\}_{i=1}^{m}\right) + \sum_{k=1}^{m}\frac{1}{m}S\left(\mathsf{a}_{jk}\right)\right) \\ &= I\left(\left\{\frac{1}{n}\right\}_{i=1}^{n}\right) + I\left(\left\{\frac{1}{m}\right\}_{i=1}^{m}\right) + \sum_{j=1}^{n}\sum_{k=1}^{m}\frac{1}{n}\frac{1}{m}S\left(\mathsf{a}_{jk}\right). \end{aligned} \tag{1.60}$$

Therefore

$$I\left(\left\{\frac{1}{nm}\right\}_{i=1}^{nm}\right) = I\left(\left\{\frac{1}{n}\right\}_{i=1}^{n}\right) + I\left(\left\{\frac{1}{m}\right\}_{i=1}^{m}\right). \tag{1.61}$$

Note that $f(n) = I\left(\left\{\frac{1}{n}\right\}_{i=1}^{n}\right)$ is a function of n only, such that $f(nm) = f(n) + f(m)$. Since I is continuous, by Cauchy's functional equation we have $f(n) = \kappa \log n$. Since the entropy is strictly concave, κ must be positive. Therefore

$$I\left(\left\{\frac{1}{n}\right\}_{i=1}^{n}\right) = \kappa \log n \tag{1.62}$$

for some $\kappa > 0$.

We now show that if the coefficients p_i are rationals, $I_n\left(\{p_i\}_{i=1}^{n}\right) = -\kappa \sum_{i=1}^{n} p_i \log p_i$. Let $\{p_i\}_{i=1}^{n}$ be rational coefficients for a convex combination. We can write them as $p_i = \frac{m_i}{m}$ where $\{m_i\}, m \in \mathbb{Z}^+$ and m is the least common denominator. Since p_i are the coefficients of a convex combination, we must have $\sum_{i=1}^{n} m_i = m$. Since m_i is a positive integer, we can write $m_i = \sum_{j=1}^{m_i} 1$. We now take m orthogonal ensembles a_{ij} where $1 \le i \le n$ and $1 \le j \le m_i$.

We have

$$\begin{aligned} S\left(\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{m}\mathsf{a}_{ij}\right) &= I\left(\left\{\frac{1}{m}\right\}_{i=1}^{m}\right)+\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{m}S\left(\mathsf{a}_{ij}\right) \\ &= \kappa\log m+\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{m}S\left(\mathsf{a}_{ij}\right) \end{aligned} \tag{1.63}$$

$$\begin{aligned} = S\left(\sum_{i=1}^{n}\frac{m_i}{m}\sum_{j=1}^{m_i}\frac{1}{m_i}\mathsf{a}_{ij}\right) &= I\left(\left\{\frac{m_i}{m}\right\}_{i=1}^{n}\right)+\sum_{i=1}^{n}\frac{m_i}{m}S\left(\sum_{j=1}^{m_i}\frac{1}{m_i}\mathsf{a}_{ij}\right) \\ &= I\left(\{p_i\}_{i=1}^{n}\right)+\sum_{i=1}^{n}\frac{m_i}{m}\left(I\left(\left\{\frac{1}{m_i}\right\}_{i=1}^{m_i}\right)+\sum_{j=1}^{m_i}\frac{1}{m_i}S\left(\mathsf{a}_{ij}\right)\right) \\ &= I\left(\{p_i\}_{i=1}^{n}\right)+\sum_{i=1}^{n}\frac{m_i}{m}I\left(\left\{\frac{1}{m_i}\right\}_{i=1}^{m_i}\right)+\sum_{i=1}^{n}\frac{m_i}{m}\sum_{j=1}^{m_i}\frac{1}{m_i}S\left(\mathsf{a}_{ij}\right) \\ &= I\left(\{p_i\}_{i=1}^{n}\right)+\sum_{i=1}^{n}p_i\kappa\log m_i+\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{m}S\left(\mathsf{a}_{ij}\right). \end{aligned} \tag{1.64}$$

Therefore

$$\begin{aligned} \kappa\log m &= I\left(\{p_i\}_{i=1}^{n}\right)+\sum_{i=1}^{n}p_i\kappa\log m_i \\ I\left(\{p_i\}_{i=1}^{n}\right) &= \kappa\log m-\sum_{i=1}^{n}p_i\kappa\log m_i=\sum_{i=1}^{n}p_i\kappa\log m-\sum_{i=1}^{n}p_i\kappa\log m_i \\ &= -\sum_{i=1}^{n}p_i\kappa\log\frac{m_i}{m}=-\kappa\sum_{i=1}^{n}p_i\log p_i. \end{aligned} \tag{1.65}$$

Lastly, let $\{p_i\}_{i=1}^{n}$ be coefficients for a convex combination, not necessarily rational. Since I is continuous and p_i can be approximated with rational values to an arbitrary level of precision, we will have $I\left(\{p_i\}_{i=1}^{n}\right)=-\kappa\sum_{i=1}^{n}p_i\log p_i$. In the case of $n=2$, we have $I(p_1,p_2)=-\kappa p_1\log p_1-\kappa p_2\log p_2$. □

Corollary 1.66. *The unit for the entropy is determined (up to the physical dimension) by the maximum of the entropy of the coefficients $I\left(\frac{1}{2},\frac{1}{2}\right)$. If the entropy is measured in bits, then $I\left(\frac{1}{2},\frac{1}{2}\right)=1$.*

Proof. A rescaling of the entropy will also rescale the entropy of the coefficients. Therefore setting the value of the maximum of I will set the arbitrary multiplicative factor. If $I\left(\frac{1}{2},\frac{1}{2}\right)=1$, then κ is equal to one and the logarithm is base two, which corresponds to the entropy measured in bits. Note that this does not fix the physical dimensions of the entropy, only the numerical value. □

Separateness and orthogonality

Recall that orthogonality is formally defined in terms of the entropy: two ensembles are orthogonal if the entropy is maximally increased during mixture. The name was chosen because

it will recover the usual notion of orthogonality for the space of distributions in both classical and quantum mechanics. Like the standard notion of orthogonality, in fact, it is irreflexive (no ensemble is orthogonal to itself) and symmetric (if a is orthogonal to b, then b is orthogonal to a). Moreover, an ensemble is not orthogonal to its own components.

Recovering the math, however, would be meaningless if we didn't also provide a stronger conceptual model. As we said, two ensembles are orthogonal if they are mutually exclusive: they have no instance in common. This means that if Amanda selects between two preparation procedures that correspond to orthogonal ensembles, Boris may only need one instance to determine which preparation was selected. A full bit of information can be transferred. This is why optimal measurements are defined across mutually exclusive, i.e. orthogonal, outcomes.

Moreover, recall that two ensembles are separate if they do not have a common component, they are not different mixtures of the same ensemble. Clearly, if two ensembles do not have instances in common, they cannot be the mixture of the same ensemble. In fact, from the axioms we posed, we are able to recover that orthogonality implies separateness. One may be tempted to conclude the converse: if two ensembles are not orthogonal, they have some instances in common and therefore we can group those common instances into a sub-ensemble of both. However, and this is the crucial problem, there is nothing that guarantees us that we can find a preparation procedure that spans only the overlap. In classical mechanics, the assumption is that we can always do it. In quantum mechanics, this does not always work and, with these concepts, it is easy to see why.

Quantum states have a lower bound on the entropy. If two pure states have an overlap, the ensemble that would corresponding to that overlap would have variability lower than a pure state: it would correspond to an entropy lower than that of a pure state. That is, we cannot produce the overlap with a reliable preparation procedure. The ensembles have instances in common but do not have an ensemble in common. This means that we have ensembles that are not mutually exclusive (i.e. not orthogonal) but do not have a common component (i.e. are separate). Understanding the difference between classical and quantum mechanics lies in understanding this case.

Proposition 1.67. *Orthogonality satisfies the following properties:*

1. *irreflexivity:* $\mathsf{a} \not\perp \mathsf{a}$
2. *symmetry:* $\mathsf{a} \perp \mathsf{b}$ *if and only if* $\mathsf{b} \perp \mathsf{a}$
3. *components are not orthogonal: if* b *is a component of* a *then* $\mathsf{a} \not\perp \mathsf{b}$
4. *orthogonality implies separateness: if* $\mathsf{a} \perp \mathsf{b}$ *then* $\mathsf{a} \pitchfork \mathsf{b}$.

Proof. For 1, let $\mathsf{a} \in \mathcal{E}$. We have $S(p\mathsf{a} + \bar{p}\mathsf{a}) = S(\mathsf{a}) < I(p, \bar{p}) + pS(\mathsf{a}) + \bar{p}S(\mathsf{a})$. Therefore a is not orthogonal to itself as it does not saturate the upper bound.

For 2, note that the upper entropy bound is symmetric in a and b.

For 3, let $\mathsf{a} = p\mathsf{b} + \bar{p}\mathsf{c}$. Since mixtures preserve orthogonality, $\mathsf{b} \perp \mathsf{a}$ if and only if $\mathsf{b} \perp \mathsf{b}$ and $\mathsf{b} \perp \mathsf{c}$. But b is not orthogonal to itself, therefore a and b are not orthogonal.

For 4, we demonstrate the contrapositive: that ensembles that are not separate are not orthogonal. Let $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ have a common component. That is, $\mathsf{a} = p\mathsf{c} + \bar{p}\mathsf{d}$ and $\mathsf{b} = \lambda\mathsf{c} + \bar{\lambda}\mathsf{e}$. Since mixtures preserve orthogonality, $\mathsf{a} \perp \mathsf{b}$ if and only if $\mathsf{a} \perp \mathsf{c}$ and $\mathsf{a} \perp \mathsf{e}$. But c is a component of a, therefore they are not orthogonal. Therefore two ensembles that have a

common component are not orthogonal. This means that if two ensembles are orthogonal they cannot have a common component and are therefore separate. □

As we did for separateness, we extend the notion of orthogonality to sets of ensembles.

Definition 1.68. *Two sets of ensembles are orthogonal if all the elements of one are orthogonal to all the elements of the other. That is, $A \perp B$ with $A, B \subseteq \mathcal{E}$ if $\mathsf{a} \perp \mathsf{b}$ for all $\mathsf{a} \in A$ and $\mathsf{b} \in B$.*

Proposition 1.69. *Let $A, B \subseteq \mathcal{E}$ be two sets of ensembles such that $A \perp B$. Then the following are true:*

1. *the two sets are separate: $A \Perp B$*
2. *their hulls are orthogonal:* $\mathrm{hull}(A) \perp \mathrm{hull}(B)$, $\mathrm{shull}(A) \perp \mathrm{shull}(B)$ *and* $\mathrm{chull}(A) \perp \mathrm{chull}(B)$

Proof. For 1, by definition $\mathsf{a} \perp \mathsf{b}$ for all $\mathsf{a} \in A$ and $\mathsf{b} \in B$. Since orthogonality implies separateness, we also have $\mathsf{a} \Perp \mathsf{b}$.

For 2, we will concentrate on the closed hull, since if two sets are orthogonal so are their subsets. Since mixtures preserve orthogonality, every finite mixture of A is orthogonal to every finite mixture of B. To show that this extends to the topological closure, let $\mathsf{a}_i \in A$ and $\mathsf{b}_j \in B$ be two sequences of finite mixtures that converge in A and B respectively. Consider $f(p, \mathsf{a}_i, \mathsf{b}_j) = S(p\mathsf{a}_i + \bar{p}\mathsf{b}_j) - pS(\mathsf{a}_i) - \bar{p}S(\mathsf{b}_j)$. Since all mixtures are orthogonal, $f(p, \mathsf{a}_i, \mathsf{b}_j) = I(p, \bar{p})$. Note that f is a continuous function, therefore the limits will also converge to $I(p, \bar{p})$. This means that every element in the closed hull of A is orthogonal to every element in the closed hull of B. □

We may want to extend the notion of separate decomposability to orthogonal decomposabilty, though it is not clear whether this will be useful or not.

Definition 1.70. *An ensemble is **orthogonally decomposable** if it can be expressed as a mixture of two orthogonal ensembles. An ensemble is **orthogonally monodecomposable** if it is both orthogonally decomposable and monodecomposable.*

Corollary 1.71. *An ensemble that is orthogonally decomposable is also separately decomposable.*

Proof. Since orthogonality implies separateness by 1.67, an orthogonal decomposition is also a separate decomposition. □

1.5 Affine combinations and vector space embedding

In this section we will see how the bounds of entropy constrain ensemble spaces to embed into vector spaces. Moreover, the ensemble space must be bounded along any direction. In other words, the existence of an entropy rules out physically pathological cases that a convex space may otherwise allow.

Affine combinations

The axiom of mixture just tells us that given two ensembles and a mixing coefficient, we can identify the mixture of the two ensembles. It does not guarantee the reverse. That is, given a final mixed ensemble, the mixing ratio and one component, it is not necessarily true that the second component is uniquely determined. Mathematically, the uniqueness of this type of decomposition is called the cancellative property of a convex space. This property is important as it can be shown that it is necessary and sufficient for the convex space to embed into a vector space (see arXiv:1105.1270).

In an ensemble space, the continuity and the strict concavity of the entropy forces the space to be cancellative. If $p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}$ for some $p \in (0,1)$, from the convex structure we can show that $p(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) + \bar{p}\mathsf{e} = p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}$ for all $p \in (0,1)$ and $\lambda \in [0,1]$. That is, if mixing e with either a or b for some mixing coefficient p yields the same result, then all non-trivial mixtures of e with any mixture of a and b will yield the same result. But this means that the entropy cannot change as we change a to b during mixture, which can only happen if they are the same ensemble.

Definition 1.72. *A convex space X is **cancellative** if $p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}$ for some $p \in (0,1)$ implies $\mathsf{a} = \mathsf{b}$.*

Theorem 1.73 (Ensemble spaces are cancellative)**.** *Let $\mathcal{E}$ be an ensemble space. Let $\mathsf{a}, \mathsf{b}, \mathsf{e} \in \mathcal{E}$ such that $p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}$ for some $p \in (0,1)$. Then $\mathsf{a} = \mathsf{b}$.*

Proof. Let $\mathsf{a}, \mathsf{b}, \mathsf{e} \in \mathcal{E}$ such that $p_0\mathsf{a} + \bar{p}_0\mathsf{e} = p_0\mathsf{b} + \bar{p}_0\mathsf{e}$ for some $p_0 \in (0,1)$.

First, we show that $p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}$ for all $p \in (0, p_0]$. In that case, since $0 < \frac{p}{p_0} \leq 1$, we have

$$p\mathsf{a} + \bar{p}\mathsf{e} = \frac{p}{p_0}(p_0\mathsf{a} + \bar{p}_0\mathsf{e}) + \overline{\left(\frac{p}{p_0}\right)}\mathsf{e} = \frac{p}{p_0}(p_0\mathsf{b} + \bar{p}_0\mathsf{e}) + \overline{\left(\frac{p}{p_0}\right)}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}. \tag{1.74}$$

Now we show that $p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}$ for all $p \in (0,1)$. Since we want to be able to expand multiple times, we want to be able to find a $p \in (0,1)$ such that

$$p\mathsf{a} + p\mathsf{a} + (1-2p)\mathsf{e} = p\mathsf{a} + \bar{p}(p_0\mathsf{a} + \bar{p}_0\mathsf{e}) = p\mathsf{a} + \bar{p}p_0\mathsf{a} + \bar{p}\bar{p}_0\mathsf{e}. \tag{1.75}$$

The coefficient of the middle term on both sides of the equality, then, has to match. That is, we want $p = \bar{p}p_0$, which means

$$\begin{aligned} p &= (1-p)p_0 = p_0 - pp_0 \\ p_0 &= p + pp_0 = (1+p_0)p \\ p &= \frac{p_0}{1+p_0} \end{aligned} \tag{1.76}$$

Note that since $p_0 > 0$ we have $p > 0$, and since the numerator is always less than the denominator $p < 1$. We have

$$\begin{aligned} 2p\mathsf{a} + \overline{2p}\mathsf{e} &= p\mathsf{a} + p\mathsf{a} + (1-2p)\mathsf{e} = p\mathsf{a} + \bar{p}(p_0\mathsf{a} + \bar{p}_0\mathsf{e}) = p\mathsf{a} + \bar{p}(p_0\mathsf{b} + \bar{p}_0\mathsf{e}) \\ &= p\mathsf{a} + p\mathsf{b} + (1-2p)\mathsf{e} = p\mathsf{b} + p\mathsf{a} + (1-2p)\mathsf{e} = p\mathsf{b} + \bar{p}(p_0\mathsf{a} + \bar{p}_0\mathsf{e}) \\ &= p\mathsf{b} + \bar{p}(p_0\mathsf{b} + \bar{p}_0\mathsf{e}) = p\mathsf{b} + p\mathsf{b} + (1-2p)\mathsf{e} \\ &= 2p\mathsf{b} + \overline{2p}\mathsf{e} \end{aligned} \tag{1.77}$$

The relationship, then, is valid for $p_1 = 2p$. Note that $p_1 > p_0$. In fact

$$\begin{aligned} p_1 &= \frac{2p_0}{1+p_0} > p_0 \\ 2p_0 &> (1+p_0)p_0 = p_0 + p_0^2 \\ p_0 &> p_0^2 \end{aligned} \tag{1.78}$$

which is true since $p_0 \in (0,1)$. Since now the relationship holds for $p_1 = \frac{2p_0}{1+p_0} > p_0$, we can repeat the process again and find that it holds for $p_2 = \frac{2p_1}{1+p_1} > p_1$ and so on. We thus have a sequence of elements between 0 and 1 and we need to determine the limit of this sequence. The only two fixed points of the expression $f(x) = \frac{2x}{1+x}$ are 0 and 1, with 0 being a repelling fixed point and 1 an attracting fixed point. Since we start with an element that is strictly between those values, the sequence will converge to 1. Therefore, since we can always find a greater mixing coefficient for which the relationship holds, combined with the previous result, $p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}$ for all $p \in (0,1)$.

Now we show that if $p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}$ for all $p \in (0,1)$, then we also have $p(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) + \bar{p}\mathsf{e} = p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e}$ for all $\lambda \in [0,1]$. We have

$$\begin{aligned} p(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) + \bar{p}\mathsf{e} &= \lambda(p\mathsf{a} + \bar{p}\mathsf{e}) + \bar{\lambda}(p\mathsf{b} + \bar{p}\mathsf{e}) = \lambda(p\mathsf{a} + \bar{p}\mathsf{e}) + \bar{\lambda}(p\mathsf{a} + \bar{p}\mathsf{e}) \\ &= p\mathsf{a} + \bar{p}\mathsf{e} = p\mathsf{b} + \bar{p}\mathsf{e} \end{aligned} \tag{1.79}$$

Lastly, we show that $S(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) = S(\mathsf{a}) = S(\mathsf{b})$ for all $\lambda \in [0,1]$. Using the entropy bounds, with $\lambda \in [0,1]$, we have

$$\begin{aligned} \lim_{p\to 1} S(p(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) + \bar{p}\mathsf{e}) &\leq \lim_{p\to 1} \left[I(p,\bar{p}) + pS(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) + \bar{p}S(\mathsf{e}) \right] = S(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) \\ \lim_{p\to 1} S(p(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) + \bar{p}\mathsf{e}) &\geq \lim_{p\to 1} \left[pS(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) + \bar{p}S(\mathsf{e}) \right] = S(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) \end{aligned} \tag{1.80}$$

and therefore

$$\lim_{p\to 1} S(p(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) + \bar{p}\mathsf{e}) = S(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}). \tag{1.81}$$

Using the above property, we also have

$$\begin{aligned} S(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) &= \lim_{p\to 1} S(p(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) + \bar{p}\mathsf{e}) = \lim_{p\to 1} S(p\mathsf{a} + \bar{p}\mathsf{e}) = S(\mathsf{a}) \\ &= \lim_{p\to 1} S(p\mathsf{b} + \bar{p}\mathsf{e}) = S(\mathsf{b}) \end{aligned} \tag{1.82}$$

Since $S(\lambda\mathsf{a}+\bar{\lambda}\mathsf{b}) = \lambda S(\lambda\mathsf{a}+\bar{\lambda}\mathsf{b}) + \bar{\lambda}S(\lambda\mathsf{a}+\bar{\lambda}\mathsf{b}) = \lambda S(\mathsf{a}) + \bar{\lambda}S(\mathsf{b})$, $\mathsf{a} = \mathsf{b}$ by strict concavity. □

The fact that the convex space is cancellative essentially allows us to "invert" a convex combination, allowing affine combinations. An affine combination is a linear combination where the coefficients sum to one, like in a convex combination, but coefficients can be negative, unlike in a convex combination. Conceptually we can "take part of an ensemble out" from another ensemble. The ability in quantum mechanics to use negative pseudo-probability, for example in Wigner functions, stems from the fact that the space of ensembles allows affine combinations because it is cancellative.

An affine combination can be translated into the equality between two convex combinations by moving all the negative coefficients on the other side of the equality. This reduces to finding the component a mixture, which, if it exists, is unique by the cancellative property.

It is important to note that while we can write affine combinations, not all affine combinations correspond to ensembles. For example, if b is not a component of a, then $\frac{3}{2}\mathsf{a} - \frac{1}{2}\mathsf{b}$ cannot possibly be an ensemble as we cannot take any "amount of b" from a.

Definition 1.83 (Affine combinations)**.** *Let $\{\mathsf{e}_i\}_{i=1}^n \subseteq \mathcal{E}$ be a finite sequence of ensembles and $\{r_i\}_{i=1}^n \subseteq \mathbb{R}$ be a finite sequence of coefficients such that $\sum_{i=1}^n r_i = 1$. The* ***affine combination*** *$\sum_{i=1}^n r_i\mathsf{e}_i$ is the ensemble $\mathsf{a} \in \mathcal{E}$, if it exists, such that $\sum_{i\in I}\frac{r_i}{r}\mathsf{e}_i = \frac{1}{r}\mathsf{a} + \sum_{i\notin I}\frac{-r_i}{r}\mathsf{e}_i$ where $I = \{i \in [1,n] \,|\, r_i \geq 0\}$ and $r = \sum_{i\in I} r_i$.*

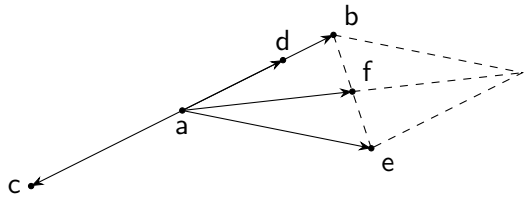
Consistency check. For the definition to work, we need to show that the affine combination can be re-expressed in terms of convex combinations. Note that the coefficients r_i are not necessarily positive, but still sum to 1. With the given definitions, I includes exactly the indexes that correspond to non-negative coefficients, and r is the sum of all those non-negative coefficients. This means that $\frac{r_i}{r} \in [0,1]$ for all $i \in I$ and $\sum_{i\in I}\frac{r_i}{r} = 1$. Therefore $\sum_{i\in I}\frac{r_i}{r}\mathsf{e}_i$ is a convex combination and it is always well-defined. Additionally, we have that $\sum_{i\in I} r_i = 1 - \sum_{i\notin I} r_i$, which means $\frac{-r_i}{r-1} \in [0,1]$ for all $i \notin I$ and $\sum_{i\in I}\frac{-r_i}{r-1} = 1$. Therefore $\sum_{i\in I}\frac{-r_i}{r-1}\mathsf{e}_i$ is a convex combination and it is always well-defined. Note that a is defined to be such that $\sum_{i\in I}\frac{r_i}{r}\mathsf{e}_i = \frac{1}{r}\mathsf{a} + \frac{r-1}{r}\sum_{i\notin I}\frac{-r_i}{r-1}\mathsf{e}_i$. If it exists, it is unique. □

Remark. In the same way that not all infinite convex combinations yield valid ensembles, not all finite or infinite affine combinations yield valid ensembles.

We still need to understand how to extend affine combination to the infinite case. The safest way is to require the two infinite mixtures $\sum_{i\in I}\frac{r_i}{r}\mathsf{e}_i$ and $\sum_{i\notin I}\frac{-r_i}{r}\mathsf{e}_i$ to converge and then require that the resulting affine problem admit a solution.

Vector space embedding

As we said before, the cancellative property is necessary and sufficient for the embedding of a convex space into a vector space. While this can be proved with a short mathematical proof, it would be an abstract construction which would not give us any insight into what is actually being represented physically. Therefore, after choosing an internal point $\mathsf{a} \in \mathcal{E}$ as an origin, we will construct a vector space from the ground up, as the space of differences between ensembles. This will also justify why, locally at every point, this space becomes the space of variations.



$$\mathsf{a} = \tfrac{1}{2}\mathsf{b} + \tfrac{1}{2}\mathsf{c} \iff 1(\mathsf{c}-\mathsf{a}) = -1(\mathsf{b}-\mathsf{a})$$

$$\mathsf{d} = \tfrac{1}{3}\mathsf{a} + \tfrac{2}{3}\mathsf{b} \iff 1(\mathsf{d}-\mathsf{a}) = \tfrac{2}{3}(\mathsf{b}-\mathsf{a})$$

$$\mathsf{f} = \tfrac{1}{2}\mathsf{b} + \tfrac{1}{2}\mathsf{e} \iff 1(\mathsf{f}-\mathsf{a}) = \tfrac{1}{2}(\mathsf{b}-\mathsf{a}) + \tfrac{1}{2}(\mathsf{e}-\mathsf{a})$$

Figure 1.3: Once a reference point is picked (i.e. a in the figure), the differences with respect to that point behave like elements of a vector space. The relationship between ensembles in terms of mixtures become vector space relationships between differences. Note that the sum is effectively a "half parallelogram law," so that the result always is a member of the convex space.

Figure 1.3 shows the motivation of the construction. We take an internal point $\mathsf{a} \in \mathcal{E}$ to be a reference, the origin. For any other ensemble b, we imagine the change from a to b as the arrow that connects the first to the second. This is the ensemble difference $(\mathsf{b}-\mathsf{a})$. The negative of that difference would find an ensemble c in the opposite direction. Note that a now sits exactly in between b and c, and therefore $\mathsf{a} = \frac{1}{2}\mathsf{b} + \frac{1}{2}\mathsf{c}$ if and only if $(\mathsf{b}-\mathsf{a}) = -(\mathsf{c}-\mathsf{a})$. That is, relationships between vectors are re-expressed as convex combinations of ensembles.

Similarly, we can now stretch or shrink differences by a factor. For example, multiplying the difference by a third will give us the ensemble that is at a third between a and b. Similarly, we can add differences. For example, $\frac{1}{2}(\mathsf{b}-\mathsf{a}) + \frac{1}{2}(\mathsf{c}-\mathsf{a})$ will give us the difference $\left(\left(\frac{1}{2}\mathsf{b} + \frac{1}{2}\mathsf{c}\right) - \mathsf{a}\right)$, the difference between a and the midpoint between b and c. Note that if we double that difference, we obtain the parallelogram law for vector addition.

Clearly, if we stretch a difference too much, we will go out of the ensemble space. Yet, since we can always shrink and then stretch, all operations defined on differences within the ensemble space can be formally defined on differences that go outside the ensemble space. Given that a is an internal point, then differences exist in every direction. This is, in a nutshell, the vector space we are constructing. This clarifies that, in the end, we are always going to be interested in sets of vectors that can be shrunk within the ensemble space.[10]

While the motivation is straightforward, the construction is complicated by the fact that there are multiple ways of expressing the same difference. In the third example in figure 1.3, $(\mathsf{b}-\mathsf{a})$ and $3(\mathsf{c}-\mathsf{a})$ correspond to the same vector. Therefore, we need to specify this equivalence class and prove that the set of equivalence classes forms a vector space. The way that the equivalence relationship is specified is so that the proofs remain short and manageable, and it is therefore mainly a technical issue.[11]

Definition 1.84 (Ensemble differences)**.** *Given an ensemble space, a difference between two ensembles represents the change required to transform one ensemble into another. Formally, an* ***ensemble difference****, noted* $r(\mathsf{b}-\mathsf{a})$*, is a triple formed by a real number*

[10] It is unclear whether this insight may help us understand what the topology of the embedding vector space is.

[11] Originally, the equivalence relationship was specified by three separate cases. While more directly justifiable, that gave 6 different possible combinations to check in every proof.

$r \in \mathbb{R}$ *and an ordered pair of ensembles* $\mathsf{a}, \mathsf{b} \in \mathcal{E}$.

Definition 1.85. *The* ***scalar multiplication*** *of a difference* $r(\mathsf{b}-\mathsf{a})$ *by a real number* $s \in \mathbb{R}$*, noted* $s(r(\mathsf{b}-\mathsf{a}))$*, is the difference* $(sr)(\mathsf{b}-\mathsf{a})$

Definition 1.86. *Two ensemble differences are* ***equivalent****, noted* $r(\mathsf{b}-\mathsf{a}) \sim s(\mathsf{c}-\mathsf{a})$*, if there exists* $k \in \mathbb{R}$ *such that*

$$\frac{1}{r+s+k}(r\mathsf{b}+s\mathsf{a}+k\mathsf{a}) = \frac{1}{r+s+k}(s\mathsf{c}+r\mathsf{a}+k\mathsf{a}).$$

Remark. Note that this is an equation between affine combinations since the coefficients can be negative. The equality can be satisfied only if both affine combinations exist.

Corollary 1.87. *If the above condition is satisfied for some* k*, then it is satisfied for all* $k \neq -(r+s)$ *for which the affine combination exists.*

Proof. Suppose $\frac{1}{r+s+k}(r\mathsf{b}+s\mathsf{a}+k\mathsf{a}) = \frac{1}{r+s+k}(s\mathsf{c}+r\mathsf{a}+k\mathsf{a})$ for some k. Let $k' \in \mathbb{R}$ be such that $\frac{1}{r+s+k'}(r\mathsf{b}+s\mathsf{a}+k'\mathsf{a})$ is a valid affine combination, which requires $r+s+k' \neq 0$. Then we have $\frac{1}{r+s+k'}(r\mathsf{b}+s\mathsf{a}+k'\mathsf{a}) = \frac{1}{r+s+k'}(r\mathsf{b}+s\mathsf{a}+k\mathsf{a}) + \frac{k'-k}{r+s+k'}\mathsf{a} = \frac{r+s+k}{r+s+k'}\frac{1}{r+s+k}(r\mathsf{b}+s\mathsf{a}+k\mathsf{a}) + \frac{k'-k}{r+s+k'}\mathsf{a} = \frac{r+s+k}{r+s+k'}\frac{1}{r+s+k}(s\mathsf{c}+r\mathsf{a}+k\mathsf{a}) + \frac{k'-k}{r+s+k'}\mathsf{a} = \frac{1}{r+s+k'}(s\mathsf{c}+r\mathsf{a}+k'\mathsf{a})$. □

Proposition 1.88. *Difference equivalence is reflexive, symmetric and transitive and therefore is an equivalence relation.*

Proof. For reflexivity, we have $\frac{1}{r+r+k}(r\mathsf{b}+r\mathsf{a}+k\mathsf{a}) = \frac{1}{r+r+k}(r\mathsf{b}+r\mathsf{a}+k\mathsf{a})$, which is satisfied for any $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ and $r \in \mathbb{R}$.

For symmetry, note that if r is switched with s and b with c, the left side becomes the right side and vice-versa.

For transitivity, suppose $r(\mathsf{b}-\mathsf{a}) \sim s(\mathsf{c}-\mathsf{a})$ and $s(\mathsf{c}-\mathsf{a}) \sim t(\mathsf{d}-\mathsf{a})$. Then for some k and l we have $\frac{1}{r+s+k}(r\mathsf{b}+s\mathsf{a}+k\mathsf{a}) = \frac{1}{r+s+k}(s\mathsf{c}+r\mathsf{a}+k\mathsf{a})$ and $\frac{1}{s+t+l}(s\mathsf{c}+t\mathsf{a}+l\mathsf{a}) = \frac{1}{s+t+l}(t\mathsf{d}+s\mathsf{a}+l\mathsf{a})$. Let $l' = r-t+k$ and therefore $k = t-r+l'$. We have $\frac{1}{r+s+k}(s\mathsf{c}+r\mathsf{a}+k\mathsf{a}) = \frac{1}{r+s+t-r+l'}(s\mathsf{c}+r\mathsf{a}+(t-r+l')\mathsf{a}) = \frac{1}{s+t+l'}(s\mathsf{c}+t\mathsf{a}+l'\mathsf{a})$. Therefore l' is such that $\frac{1}{s+t+l'}(s\mathsf{c}+t\mathsf{a}+l'\mathsf{a})$ is a valid affine combination, which means $\frac{1}{s+t+l'}(s\mathsf{c}+t\mathsf{a}+l'\mathsf{a}) = \frac{1}{s+t+l'}(t\mathsf{d}+s\mathsf{a}+l'\mathsf{a})$. Now let $k' = l'+s-r$ which gives $l' = k'+r-s$ and $k' = k+s-t$. We have $\frac{1}{r+t+k'}(r\mathsf{b}+t\mathsf{a}+k'\mathsf{a}) = \frac{1}{r+s+k}(r\mathsf{b}+s\mathsf{a}+k\mathsf{a}) = \frac{1}{r+s+k}(s\mathsf{c}+r\mathsf{a}+k\mathsf{a}) = \frac{1}{s+t+l'}(t\mathsf{d}+s\mathsf{a}+l'\mathsf{a}) = \frac{1}{r+t+k'}(t\mathsf{d}+r\mathsf{a}+k'\mathsf{a})$. Therefore $r(\mathsf{b}-\mathsf{a}) \sim t(\mathsf{d}-\mathsf{a})$. □

Proposition 1.89. *The ensemble difference equivalence relation satisfies the following:*

1. *if* $r(\mathsf{b}-\mathsf{a}) \sim s(\mathsf{c}-\mathsf{a})$ *with* $r \neq 0$ *and* $s \neq 0$*, then* a*,* b *and* c *are on the same line*
2. *if* $r \neq 0$*,* $[r(\mathsf{b}-\mathsf{a})] = \left\{(r+j)\left(\left(\frac{r}{r+j}\mathsf{b}+\frac{j}{r+j}\mathsf{a}\right)-\mathsf{a}\right)\right\}$ *for all* $j \in \mathbb{R}$ *such that the affine combination exists*
3. $[0(\mathsf{b}-\mathsf{a})] = \{0\,(\mathsf{e}-\mathsf{a}) \mid \mathsf{e} \in \mathcal{E}\} \cup \{r\,(\mathsf{a}-\mathsf{a}) \mid r \in \mathbb{R}\}$
4. *given* $p \in (0,1]$*,* $r(\mathsf{b}-\mathsf{a}) \sim \frac{r}{p}((p\mathsf{b}+\bar{p}\mathsf{a})-\mathsf{a})$
5. $r(\mathsf{b}-\mathsf{a}) \sim s(\mathsf{c}-\mathsf{a})$ *implies* $(rt)(\mathsf{b}-\mathsf{a}) \sim (st)(\mathsf{c}-\mathsf{a})$ *and therefore scalar multiplication maps equivalence classes to equivalence classes*

Proof. For 1, the condition for equivalence imposes that an affine combination of b and a equals an affine combination of c and a. Equivalently, this says that they are expressible as affine combinations of each other, that they are on the same line.

For 2, let us first verify that $r(\mathsf{b}-\mathsf{a}) \sim (r+j)\left(\left(\frac{r}{r+j}\mathsf{b}+\frac{j}{r+j}\mathsf{a}\right)-\mathsf{a}\right)$. We have $\frac{1}{r+r+j+k}(r\mathsf{b}+(r+j)\mathsf{a}+k\mathsf{a}) = \frac{1}{r+r+j+k}\left((r+j)\left(\frac{r}{r+j}\mathsf{b}+\frac{j}{r+j}\mathsf{a}\right)+r\mathsf{a}+k\mathsf{a}\right) = \frac{1}{r+r+j+k}\left(r\mathsf{b}+j\mathsf{a}+r\mathsf{a}+k\mathsf{a}\right)$, which means the two differences are equivalent. Note that all ensembles that are on the same line of a and b can, except for a, be expressed as $\frac{r}{r+j}\mathsf{b}+\frac{j}{r+j}\mathsf{a}$. Since, by 1, being on the same line is a necessary condition for equivalence, the expression spans the entire equivalence class.

For 3, suppose $r(\mathsf{c}-\mathsf{a}) \sim 0(\mathsf{b}-\mathsf{a})$. Then $\frac{1}{r+0+k}(r\mathsf{c}+0\mathsf{a}+k\mathsf{a}) = \frac{1}{r+0+k}(0\mathsf{b}+r\mathsf{a}+k\mathsf{a})$. If $r=0$ we have $\mathsf{a}=\mathsf{a}$ which is satisfied for all $\mathsf{c}\in\mathcal{E}$. If $r\neq 0$ we have $\frac{1}{r+k}(r\mathsf{c}+k\mathsf{a})=\mathsf{a}$ which is satisfied only if $\mathsf{c}=\mathsf{a}$.

For 4, let $\frac{r}{p} = r+j$. We have $j = \frac{r}{p}-r = \frac{r(1-p)}{p} = (r+j)\bar{p}$. Therefore $\frac{r}{r+j} = r\frac{p}{r} = p$ and $\frac{j}{r+j}=\bar{p}$. This matches the coefficients in the equivalence class.

For 5, we have

$$\begin{aligned}
\frac{1}{r+s+k}(r\mathsf{b}+s\mathsf{a}+k\mathsf{a}) &= \frac{1}{r+s+k}(s\mathsf{c}+r\mathsf{a}+k\mathsf{a}) \\
\frac{1}{rt+st+kt}(rt\mathsf{b}+st\mathsf{a}+kt\mathsf{a}) &= \frac{1}{rt+st+kt}(st\mathsf{c}+rt\mathsf{a}+kt\mathsf{a}) \\
\frac{1}{rt+st+k'}(rt\mathsf{b}+st\mathsf{a}+k'\mathsf{a}) &= \frac{1}{rt+st+k'}(st\mathsf{c}+rt\mathsf{a}+k'\mathsf{a})
\end{aligned} \tag{1.90}$$

which means $(rt)(\mathsf{b}-\mathsf{a}) \sim (st)(\mathsf{c}-\mathsf{a})$. □

Definition 1.91. *The **addition** between two equivalence classes of differences is defined by*

$$[r(\mathsf{b}-\mathsf{a})]+[s(\mathsf{c}-\mathsf{a})] = \left[(r+s+k)\left(\left(\frac{r}{r+s+k}\mathsf{b}+\frac{s}{r+s+k}\mathsf{c}+\frac{k}{r+s+k}\mathsf{a}\right)-\mathsf{a}\right)\right]$$

where $r+s+k\neq 0$.

Consistency check. We need to show that the definition does not depend on the chosen representatives of the classes on the left hand side. We first check the case when both classes are of the form $[0(\mathsf{b}-\mathsf{a})]$. Suppose we have two elements of said equivalence class. Either the coefficient is zero or the two ensembles are the same. In either case, the elements on the right side of the definition of the addition will be affine combinations of a only. Therefore the result of the addition of $[0(\mathsf{b}-\mathsf{a})]$ with itself gives us $[0(\mathsf{b}-\mathsf{a})]$. Now suppose we sum an equivalence class $[r(\mathsf{b}-\mathsf{a})]$ with $[0(\mathsf{c}-\mathsf{a})]$. Note that either $s=0$ or $\mathsf{c}=\mathsf{a}$. In either case, the ensemble in the result is an affine combination of only b and a for every k. The result, then, spans exactly the equivalence class $[r(\mathsf{b}-\mathsf{a})]$. Lastly, suppose we are summing two equivalence classes different from $[0(\mathsf{b}-\mathsf{a})]$. The ratio between b and c remains the same for all k, therefore the ensembles in the result span the line between a and an affine combination of b and c, which means all elements fall in the same equivalence class. □

Corollary 1.92. *Let* $\mathsf{a} \in \mathcal{E}$ *be an ensemble and let* $V = \{[r(\mathsf{b}-\mathsf{a})]\}$ *be the set of equivalence classes of ensemble differences from* a*. Then the addition of ensemble difference over* V *is a commutative monoid. If* a *is an internal point, the addition is an abelian group.*

Proof. The operation is commutative as the definition of addition is symmetric over its arguments. The operation is associative as further sums of equivalence classes turn into additions of the real coefficients and extensions of affine combinations both of which are associative. The zero equivalence class $[0(\mathsf{b}-\mathsf{a})]$ is the identity element, as shown in the previous proof. The addition is a commutative monoid.

Now let a be an internal point and let $[r(\mathsf{b}-\mathsf{a})]$ be an equivalence class. Since a is an internal point, we can find $\mathsf{c} \in \mathcal{E}$ such that $\mathsf{a} = p\mathsf{b} + \bar{p}\mathsf{c}$. Consider $\left[r\frac{\bar{p}}{p}(\mathsf{c}-\mathsf{a})\right]$. Note that $r + r\frac{\bar{p}}{p} = \frac{r}{p}(p+\bar{p}) = \frac{r}{p}$. The addition is given by

$$\begin{aligned}
&\left[\left(r + r\frac{\bar{p}}{p} + k\right)\left(\left(\frac{r}{r + r\frac{\bar{p}}{p} + k}\mathsf{b} + \frac{r\frac{\bar{p}}{p}}{r + r\frac{\bar{p}}{p} + k}\mathsf{c} + \frac{k}{r + r\frac{\bar{p}}{p} + k}\mathsf{a}\right) - \mathsf{a}\right)\right] \\
&= \left[\left(\frac{r}{p} + k\right)\left(\left(\frac{\frac{r}{p}}{\frac{r}{p} + k}(p\mathsf{b} + \bar{p}\mathsf{c}) + \frac{k}{\frac{r}{p} + k}\mathsf{a}\right) - \mathsf{a}\right)\right] \\
&= \left[\left(\frac{r}{p} + k\right)\left(\left(\frac{\frac{r}{p}}{\frac{r}{p} + k}\mathsf{a} + \frac{k}{\frac{r}{p} + k}\mathsf{a}\right) - \mathsf{a}\right)\right] \\
&= [0(\mathsf{b}-\mathsf{a})].
\end{aligned} \tag{1.93}$$

This means that $\left[r\frac{\bar{p}}{p}(\mathsf{c}-\mathsf{a})\right]$ is the inverse of $[r(\mathsf{b}-\mathsf{a})]$ and that the addition is an abelian group. □

Theorem 1.94 (Differences form a vector space)**.** *Let* $\mathsf{a} \in \mathcal{E}$ *be an internal point and let* $V = \{[r(\mathsf{b}-\mathsf{a})]\}$ *be the set of equivalence classes of ensemble differences from* a*. Then* V *is a vector space under scalar multiplication and addition.*

Proof. Scalar multiplication is compatible with the multiplication between scalars since $r(s(t(\mathsf{b}-\mathsf{a}))) = r((st)(\mathsf{b}-\mathsf{a})) = (rst)(\mathsf{b}-\mathsf{a}) = (rs)(t(\mathsf{b}-\mathsf{a}))$. The identity of the scalars is the identity of scalar multiplication since $1(r(\mathsf{b}-\mathsf{a})) = r(\mathsf{b}-\mathsf{a})$. Scalar multiplication is

distributive with respect to vector addition:

$$\begin{aligned}
&t\left([r(\mathsf{b}-\mathsf{a})]+[s(\mathsf{c}-\mathsf{a})]\right)\\
&=t\left[(r+s+k)\left(\left(\frac{r}{r+s+k}\mathsf{b}+\frac{s}{r+s+k}\mathsf{c}+\frac{k}{r+s+k}\mathsf{a}\right)-\mathsf{a}\right)\right]\\
&=\left[t(r+s+k)\left(\left(\frac{r}{r+s+k}\mathsf{b}+\frac{s}{r+s+k}\mathsf{c}+\frac{k}{r+s+k}\mathsf{a}\right)-\mathsf{a}\right)\right]\\
&=\left[(tr+ts+tk)\left(\left(\frac{tr}{tr+ts+tk}\mathsf{b}+\frac{ts}{tr+ts+tk}\mathsf{c}+\frac{tk}{tr+ts+tk}\mathsf{a}\right)-\mathsf{a}\right)\right]\\
&=\left[(tr+ts+k')\left(\left(\frac{tr}{tr+ts+k'}\mathsf{b}+\frac{ts}{tr+ts+k'}\mathsf{c}+\frac{k'}{tr+ts+k'}\mathsf{a}\right)-\mathsf{a}\right)\right]\\
&[tr(\mathsf{b}-\mathsf{a})]+[ts(\mathsf{c}-\mathsf{a})]\,.
\end{aligned}\tag{1.95}$$

Scalar multiplication is distributive with respect to scalar addition

$$\begin{aligned}
&s\,[r(\mathsf{b}-\mathsf{a})]+t\,[r(\mathsf{b}-\mathsf{a})]\\
&[sr(\mathsf{b}-\mathsf{a})]+[tr(\mathsf{b}-\mathsf{a})]\\
&=\left[(sr+tr+k)\left(\left(\frac{sr}{sr+tr+k}\mathsf{b}+\frac{tr}{sr+tr+k}\mathsf{b}+\frac{k}{sr+tr+k}\mathsf{a}\right)-\mathsf{a}\right)\right]\\
&=\left[(sr+tr+k)\left(\left(\frac{sr+tr}{sr+tr+k}\mathsf{b}+\frac{k}{sr+tr+k}\mathsf{a}\right)-\mathsf{a}\right)\right]\\
&=[((s+t)r)(\mathsf{b}-\mathsf{a})]\\
&=(s+t)[r(\mathsf{b}-\mathsf{a})].
\end{aligned}\tag{1.96}$$

Therefore V with addition and scalar multiplication satisfies the definition of a vector space. □

Definition 1.97. *Given an internal point* a*, the* ***natural embedding*** *of* $\mathcal{E}$ *into* V_{a} *is the map* $\iota_{\mathsf{a}}:\mathcal{E}\hookrightarrow V_{\mathsf{a}}$*, defined as* $\iota_{\mathsf{a}}(\mathsf{e})\to[1(\mathsf{e}-\mathsf{a})]$*, that maps each ensemble to its difference from* a*.*

Corollary 1.98. *The natural embedding of* $\mathcal{E}$ *into* V_{a} *preserves affine combinations. That is,* $\iota\left(\sum_{i=1}^{n} r_i\mathsf{e}_i\right)=\sum_{i=1}^{n} r_i\iota(\mathsf{e}_i)$*.*

Proof. Let $r+s=1$ and $\mathsf{b},\mathsf{c}\in\mathcal{E}$ such that $r\mathsf{b}+s\mathsf{c}\in\mathcal{E}$. We have

$$\begin{aligned}
r[1(\mathsf{b}-\mathsf{a})]+s[1(\mathsf{c}-\mathsf{a})]&=[r(\mathsf{b}-\mathsf{a})]+[s(\mathsf{c}-\mathsf{a})]\\
&=\left[(r+s+k)\left(\left(\frac{r}{r+s+k}\mathsf{b}+\frac{s}{r+s+k}\mathsf{c}+\frac{k}{r+s+k}\mathsf{a}\right)-\mathsf{a}\right)\right]\\
&=\left[(1+k)\left(\left(\frac{r}{1+k}\mathsf{b}+\frac{s}{1+k}\mathsf{c}+\frac{k}{1+k}\mathsf{a}\right)-\mathsf{a}\right)\right]\\
&=\left[1\left(\left(\frac{r}{1}\mathsf{b}+\frac{s}{1}\mathsf{c}\right)-\mathsf{a}\right)\right]\\
&=[1\left((r\mathsf{b}+s\mathsf{c})-\mathsf{a}\right)]\,.
\end{aligned}\tag{1.99}$$

If we apply the above recursively, we have $\sum_{i=1}^{n} r_i[1(\mathsf{e}_i - \mathsf{a})] = [1\,((\sum_{i=1}^{n} r_i \mathsf{e}_i) - \mathsf{a})]$. Therefore $\sum_{i=1}^{n} r_i \iota(\mathsf{e}_i) = \iota\,(\sum_{i=1}^{n} r_i \mathsf{e}_i)$. □

Proposition 1.100. *Let $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ be two internal points. Then $r(\mathsf{c} - \mathsf{b}) \mapsto r(\mathsf{c} - \mathsf{a}) - r(\mathsf{b} - \mathsf{a})$ is an isomorphism between the corresponding vector spaces.*

Proof. Note that the natural embeddings of $\mathcal{E}$ in V_{a} and V_{b} are invertible onto their images, and therefore they give an affine map between V_{a} and V_{b}. That is, $r(\mathsf{c}-\mathsf{b}) \mapsto r(\mathsf{c}-\mathsf{a})$ is an affine map. Moreover, $r(\mathsf{c} - \mathsf{b}) \mapsto r(\mathsf{c} - \mathsf{a}) + v$ will be an affine map for any $v \in V_{\mathsf{a}}$. An affine map is linear if and only if it maps the zero vector to the zero vector. Note that $r(\mathsf{b} - \mathsf{b})$ is the zero vector for V_{b}, which means the above map is affine if and only if $r(\mathsf{b} - \mathsf{b}) \mapsto r(\mathsf{a} - \mathsf{a}) \sim r(\mathsf{b} - \mathsf{a}) - r(\mathsf{b} - \mathsf{a})$ and therefore $v = -r(\mathsf{b} - \mathsf{a})$. □

We leave open whether the ensemble space embeds continuously in a topological vector space, and whether the topological vector space would have to be a locally convex. The first step would be to check whether this is true in the finite-dimensional case.

Self-mixtures

As an example of a convex space that is non-cancellative, does not embed in a vector space, and therefore is ruled out as an ensemble space, we consider one in which the convex combination of two different elements returns the first. We show that this can be made to satisfy the axioms of ensemble and mixture, and therefore it is really the axiom of entropy that rules it out.

Example 1.101 (Self-mixture). Let $\mathcal{E} = \{\mathsf{a}, \mathsf{b}\}$ be a set endowed with the convex structure that satisfies identity, idempotence, commutativity and such that $p\mathsf{a} + \bar{p}\mathsf{b}$ equals a if $p = 1$ and b otherwise. Identity, idempotence and commutativity are satisfied by construction. For associativity, note that the final result of multiple mixtures is a if and only if all elements of the mixtures are a. The order does not matter, and therefore associativity is satisfied. Therefore such structure satisfies at least the axiom of mixture without the requirement of continuity.

For continuity, we need to look at the inverse image under the mixing operation. Note that $+^{-1}(\varnothing) = \varnothing$ and $+^{-1}(\mathcal{E}) = [0,1] \times \mathcal{E} \times \mathcal{E}$. Since $\varnothing$ and $\mathcal{E}$ must be in any topology of $\mathcal{E}$, the continuity condition is satisfied for these sets. Next, we have $+^{-1}(\{\mathsf{a}\}) = [0,1] \times \{\mathsf{a}\} \times \{\mathsf{a}\} \cup [1] \times \{\mathsf{a}\} \times \{\mathsf{b}\} \cup [0] \times \{\mathsf{b}\} \times \{\mathsf{a}\}$. Note that $[0]$ and $[1]$ are not open sets, therefore $\{\mathsf{a}\}$ cannot be in the topology of $\mathcal{E}$ or mixing would not be continuous. Finally, $+^{-1}(\{\mathsf{b}\}) = [0,1] \times \{\mathsf{b}\} \times \{\mathsf{b}\} \cup [0,1) \times \{\mathsf{a}\} \times \{\mathsf{b}\} \cup (0,1] \times \{\mathsf{b}\} \times \{\mathsf{a}\} = [0,1) \times \mathcal{E} \times \{\mathsf{b}\} \cup (0,1] \times \{\mathsf{b}\} \times \mathcal{E}$. Since $(0,1]$ and $[0,1)$ are open subsets of $[0,1]$, $\{\mathsf{b}\}$ can be an open set. Since the topology must be T_0, we must include at least $\{\mathsf{b}\}$ and therefore $\mathsf{T}_{\mathcal{E}} = \{\varnothing, \{\mathsf{b}\}, \{\mathsf{a}, \mathsf{b}\}\}$ is a T_0 second countable topology for which mixing is continuous. This means that the example satisfies both the axioms of ensemble and of mixture.

It is therefore the entropy that rules out this case. If the entropy of the two ensembles is different, it would jump from $S(\mathsf{a})$ directly to $S(\mathsf{b})$ for an infinitesimal mixture which violates the upper bound. If $S(\mathsf{a})$ equals $S(\mathsf{b})$, the entropy of the mixture of a and b is also the same, so the two ensembles cannot possibly be different by strict concavity.

Intuitively, this tells us that we cannot have a mixture of two different elements that happens to be equal to one of the original elements.

Corollary 1.102. *Let* $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ *such that* $p\mathsf{a} + \bar{p}\mathsf{b} = \mathsf{b}$ *for some* $p \in (0, 1]$. *Then* $\mathsf{a} = \mathsf{b}$.

Proof. We have $p\mathsf{a} + \bar{p}\mathsf{b} = \mathsf{b} = p\mathsf{b} + \bar{p}\mathsf{b}$ for some $p \in (0, 1]$. Therefore $\mathsf{a} = \mathsf{b}$. □

Boundedness of lines

Another constraint that the entropy imposes is that the ensemble space is bounded along every direction. That is, if we take two ensembles, these will identify an affine line in the embedding vector space that will, at some point, exit the ensemble space in both directions.

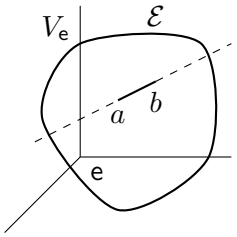


Figure 1.4: $\mathcal{E}$ is the ensemble space and V_{e} is the embedding vector space with origin e. Any affine line, like the dashed one in the picture, will intersect $\mathcal{E}$ over a finite range limited by the two elements a and b of the vector space (but not necessarily of the ensemble space).

The proof proceeds as follows. Suppose we take three points and assign them an entropy value consistent with the entropic bounds: can we always pick an entropy value that satisfies the bounds for any subsequent point in a given direction? The answer is negative and figure 1.5 represents the problem pictorially. On the horizontal axis we can imagine all the ensembles over a line, and on the vertical axis their entropy. We fix the entropy for the origin, the blue and the red point. What are the values for the next point, the green one, that satisfy the entropy bounds? Because of strict concavity, the green point must remain below the blue line. The green line represents the upper bound for the blue point, which means the green point must be above the purple line. If we go far enough, the blue and purple line meet, leaving no possible value for the entropy.

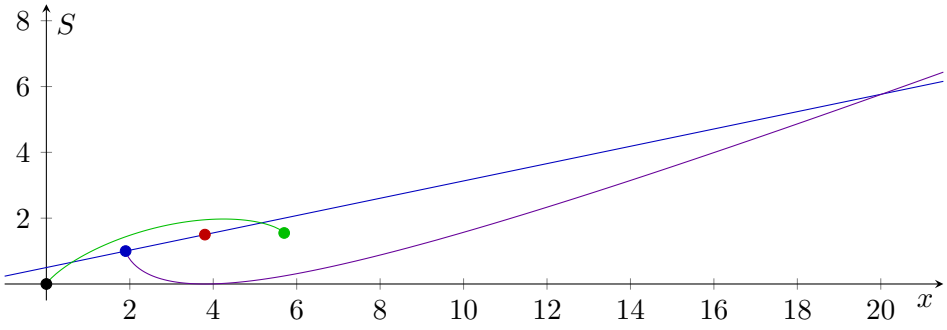


Figure 1.5: Visual representation of the bounds. See proof for details.

Definition 1.103. *A **line** $A \subseteq \mathcal{E}$ is a convex subset such that for any three elements one can be expressed as a mixture of the other two. That is, for all distinct $\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3 \in A$ there exists a permutation $\sigma : \{1,2,3\} \to \{1,2,3\}$ and $p \in (0,1)$ such that $\mathsf{e}_{\sigma(1)} = p\mathsf{e}_{\sigma(2)} + \bar{p}\mathsf{e}_{\sigma(3)}$.*

Proposition 1.104. *Given two distinct $\mathsf{a}, \mathsf{b} \in \mathcal{E}$, there is only one line that contains them.*

Proof. Since an ensemble space embeds into a vector space, given two distinct ensembles there is only one affine line that connects them. A line is the intersection of the affine line with the ensemble space. □

Theorem 1.105 (Lines are bounded)**.** *Let $A \subseteq \mathcal{E}$ be a line. Then we can find a bounded interval $V \subset \mathbb{R}$ and an invertible function $f : A \to V$ such that $f(p\mathsf{a} + \bar{p}\mathsf{b}) = pf(\mathsf{a}) + \bar{p}f(\mathsf{b})$ for all $\mathsf{a}, \mathsf{b} \in A$.*

Proof. Let $A \subseteq \mathcal{E}$ be a line. Pick $\mathsf{e}_0, \mathsf{e}_1 \in A$. For any $\mathsf{a} \in A$, we can write it as an affine combination of e_0 and e_1. That is, $\mathsf{a} = \bar{x}\mathsf{e}_0 + x\mathsf{e}_1$ where $x \in \mathbb{R}$. Note that a uniquely determines x. Therefore we can define $f(\mathsf{a}) \mapsto x$, and it will be an invertible function.

We can verify that, given $\mathsf{a}, \mathsf{b} \in A$, we have

$$\begin{aligned} f(p\mathsf{a} + \bar{p}\mathsf{b}) &= f(p\bar{x}_a\mathsf{e}_0 + px_a\mathsf{e}_1 + \bar{p}\bar{x}_b\mathsf{e}_0 + \bar{p}x_b\mathsf{e}_1) = f((p\bar{x}_a + \bar{p}\bar{x}_b)\mathsf{e}_0 + (px_a + \bar{p}x_b)\mathsf{e}_1) \\ &= px_a + \bar{p}x_b = pf(\mathsf{a}) + \bar{p}f(\mathsf{b}). \end{aligned} \tag{1.106}$$

We are going to show that the image $V = f(A)$ must be a bounded set, or it would eventually violate the entropy bounds. Given a value $x \in V$, there will be an entropy value $S(f^{-1}(x))$ which we can write as a function of the real value $S(x)$. This function will need to satisfy continuity, strict concavity and the upper variability bound.

We are going to show that the function $S(x)$ cannot extend to plus infinity. The same argument can be applied by symmetry for minus infinity. We are going to assume that $S(0) = 0$ without loss of generality. If S is not defined at zero, or if the value is different, we can apply a translation on the argument or on the value, which will not affect the concavity of the function.

Let $0 < a < b \in V$ be two distinct values, with $S(a)$ and $S(b)$ their respective entropies. We are going to show that if we pick a $c > b$ sufficiently large, we are not going to find a value for $S(c)$ that satisfies the bounds. In the picture, we have the three points, a in blue, b in red, c in green. The horizontal axis represents the value, the position of the ensemble along the affine line, while the vertical axis represents the entropy. Consider the points a, b and c. Since the entropy is strictly concave, c must be placed under the blue line which represents an upper bound on $S(c)$. Now consider 0, a and c. Since a is a mixture of 0 and c, $S(a)$ will need to satisfy the upper bound given by $S(0)$ and $S(c)$. In the diagram, the green line represents the upper bound on $S(a)$ for the specific choice of c and $S(c)$. Since a and $S(a)$ are fixed, this puts a lower bound on $S(c)$, which is represented by the purple curve. That is, the purple curve represents the minimum value we have to assign to $S(c)$ such that $S(a)$ still satisfies the upper bound between 0 and c. As the picture shows, the two bounds meet at some point and cannot both be satisfied.

From strict concavity, noting that $\frac{c-b}{c-a} + \frac{b-a}{c-a} = 1$, we have:

$$\begin{aligned} S(b) &= S\left(\frac{c-b}{c-a}a + \frac{b-a}{c-a}c\right) > \frac{c-b}{c-a}S(a) + \frac{b-a}{c-a}S(c) \\ (b-a)S(c) &< (c-a)S(b) - (c-b)S(a) = (c-a)S(b) - (c-a)S(a) + (b-a)S(a) \\ S(c) &< \frac{S(b)-S(a)}{(b-a)}(c-a) + S(a) \end{aligned} \tag{1.107}$$

From the upper bound, noting that $\frac{c-a}{c} + \frac{a}{c} = 1$ and recalling we assumed $S(0) = 0$, we have:

$$\begin{aligned} S(a) &= S\left(\frac{c-a}{c}0 + \frac{a}{c}c\right) \leq I\left(\frac{c-a}{c}, \frac{a}{c}\right) + \frac{c-a}{c}S(0) + \frac{a}{c}S(c) \\ \frac{a}{c}S(c) &\geq S(a) - I\left(\frac{c-a}{c}, \frac{a}{c}\right) \\ S(c) &\geq \frac{c}{a}\left[S(a) - I\left(\frac{c-a}{c}, \frac{a}{c}\right)\right] \end{aligned} \tag{1.108}$$

Combining the bounds, we have:

$$\begin{aligned} \frac{S(b)-S(a)}{(b-a)}(c-a) + S(a) &> \frac{c}{a}\left[S(a) - I\left(\frac{c-a}{c}, \frac{a}{c}\right)\right] \\ (c-a)S(b) - (c-a)S(a) + (b-a)S(a) &> \frac{c(b-a)}{a}S(a) - \frac{c(b-a)}{a}I\left(\frac{c-a}{c}, \frac{a}{c}\right) \\ a(c-a)S(b) - a(c-b)S(a) &> c(b-a)S(a) - c(b-a)I\left(\frac{c-a}{c}, \frac{a}{c}\right) \\ a(c-a)S(b) + (-ac+ab-bc+ac)S(a) &> -c(b-a)I\left(\frac{c-a}{c}, \frac{a}{c}\right) \\ a(c-a)S(b) - b(c-a)S(a) &> -c(b-a)I\left(\frac{c-a}{c}, \frac{a}{c}\right) \\ aS(b) - bS(a) &> -\frac{c(b-a)}{(c-a)}I\left(\frac{c-a}{c}, \frac{a}{c}\right) \end{aligned} \tag{1.109}$$

Since $b > a$, $c > a$ and $I(p, \bar{p}) > 0$ for all $p \in (0, 1)$, the right hand side of the inequality is always negative. As c increases, the right hand side will go to zero, since $\lim_{c\to\infty} \frac{c(b-a)}{c-a} = b - a$ and $\lim_{c\to\infty} I\left(\frac{c-a}{c}, \frac{a}{c}\right) = I(1, 0) = 0$. The left hand side is a constant. If the constant is positive, the inequality is always satisfied. If it is negative, it will not be satisfied for all c.

From strict concavity, noting that $\frac{b-a}{b} + \frac{a}{b} = 1$, we have

$$\begin{aligned} S(a) &= S\left(\frac{b-a}{b}0 + \frac{a}{b}b\right) > \frac{b-a}{b}S(0) + \frac{a}{b}S(b) = \frac{a}{b}S(b) \\ bS(a) &> aS(b) \\ 0 &> aS(b) - bS(a) \end{aligned} \tag{1.110}$$

This shows that the left hand side of the previous inequality is negative, and therefore the bounds cannot be satisfied over the whole $\mathbb{R}$.

This means that $V = f(A)$ must be bounded and therefore every line is a segment as embedded in the vector space. □

Remark. Note that this does not mean that, along each direction, the line is closed in the vector space. That is, it may not include the extreme points in the convex space. An open bounded interval, in fact, is still a convex space and we would be able to define an entropy on it.

The fact that the ensemble space is directionally bounded gives us an intuitive property that we would, mistakenly, always think to be true. Suppose we have an ensemble e and a sequence of ensembles $\mathsf{a}_i \in L$ on some line L that contains e. Then $p_i\mathsf{e} + \bar{p}_i\mathsf{a}_i \to \mathsf{e}$ if $p_i \to 1$. This does not work in a generic convex space if directions are unbounded. For example, suppose $\mathsf{e} = \mathbb{R}$, which is a convex set but not directionally bounded. Let $\mathsf{e} = 0$, $\mathsf{a}_i = i$ and $p_i = 1 - \frac{1}{i}$. Then $\left(1 - \frac{1}{i}\right)0 + \frac{1}{i}i = 1$. Therefore $p_i\mathsf{e} + \bar{p}_i\mathsf{a}_i \nrightarrow \mathsf{e}$ even though $p_i \to 1$. It is, again, the entropy that guarantees that this intuitive property is satisfied.

Proposition 1.111. *Let $L \subseteq \mathcal{E}$ be a line. Let $\mathsf{e} \in L$ and $\mathsf{a}_i \in L$. Then $p_i\mathsf{e} + \bar{p}_i\mathsf{a}_i \to \mathsf{e}$ if $p_i \to 1$.*

Proof. Pick $\mathsf{a} \in L$ such that $\mathsf{a} \neq \mathsf{e}$. Then every a_i can be expressed as an affine combination $\mathsf{a}_i = r_i\mathsf{e} + \bar{r}_i\mathsf{a}$. We have $p_i\mathsf{e} + \bar{p}_i\mathsf{a}_i = p_i\mathsf{e} + \bar{p}_i\left(r_i\mathsf{e} + \bar{r}_i\mathsf{a}\right) = \left(p_i + \bar{p}_i r_i\right)\mathsf{e} + \bar{p}_i\bar{r}_i\mathsf{a}$. Since the ensemble space is directionally bounded, the sequence r_i is bounded from below and above. This means that $\bar{p}_i \to 0$ implies $p_i + \bar{p}_i r_i \to 1$ and $\bar{p}_i\bar{r}_i \to 0$. Therefore by continuity $\left(p_i + \bar{p}_i r_i\right)\mathsf{e} + \bar{p}_i\bar{r}_i\mathsf{a} \to 1\mathsf{e} + 0\mathsf{a} = \mathsf{e}$. □

It is an open question whether this property is guaranteed for a generic sequence of ensembles a_i, not necessarily on a line. At least, it should be extensible to the finite-dimensional case.

Conjecture 1.112. *Let $\mathsf{e} \in \mathcal{E}$ and $\mathsf{a}_i \in \mathcal{E}$. Then $p_i\mathsf{e} + \bar{p}_i\mathsf{a}_i \to \mathsf{e}$ if $p_i \to 1$.*

As discussed previously, it may be that the boundedness provided by the entropy forces all submixtures of convergent infinite mixtures to converge.

Conjecture 1.113. *Let $\mathsf{a} = \sum_{i=1}^{\infty} p_i\mathsf{e}_i$. Then, given $I \subseteq \mathbb{N}$, $\sum_{i\in I} \frac{p_i}{p_I}\mathsf{e}_i$ converges where $p_I = \sum_{i\in I} p_i$.*

Conjecture 1.114. *The set of all internal points $I_{\mathcal{E}}$ is closed under infinite convex combinations.*

1.6 Entropic geometry

In this section we will see how the entropy imposes a geometric structure on the ensemble space. The core idea is that, since the entropy is strictly concave, its Hessian is negative definite. The negation is therefore a positive definite, real-valued function of two variations and plays the role of a metric tensor.

Mixing entropy as pseudo-distance

The first observation is that the entropy can be used to define a pseudo-distance. If we mix two ensembles, in fact, the average entropy cannot decrease and will stay the same if the two components of the mixture are the same ensemble. The increase, then, is zero if the two

ensembles are equal, greater than zero if not, with the maximum if they are orthogonal. This means that the increase of entropy during mixing can be used to characterize how different two ensembles are.

We define the mixing entropy as the increase in entropy for the equal mixture of two states. This satisfies all axioms for a distance, except the triangle inequality. The mixing entropy recovers the Jensen-Shannon divergence in both the classical and quantum case, and can therefore be seen as its generalization.

Definition 1.115. *Given two ensembles* $\mathsf{a}, \mathsf{b} \in \mathcal{E}$*, the* ***mixing entropy****, also called Jensen-Shannon divergence, is the increase in entropy associated to their equal mixture. That is:*

$$MS(\mathsf{a}, \mathsf{b}) = S\left(\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}\right) - \left(\frac{1}{2}S(\mathsf{a}) + \frac{1}{2}S(\mathsf{b})\right).$$

Proposition 1.116. *The mixing entropy* $MS(\mathsf{a}, \mathsf{b})$ *satisfies the following:*

1. ***non-negativity****:* $MS(\mathsf{a}, \mathsf{b}) \geq 0$
2. ***identity of indiscernibles****:* $MS(\mathsf{a}, \mathsf{b}) = 0 \iff \mathsf{a} = \mathsf{b}$
3. ***unit boundedness****:* $MS(\mathsf{a}, \mathsf{b}) \leq 1$
4. ***maximality of orthogonals****:* $MS(\mathsf{a}, \mathsf{b}) = 1 \iff \mathsf{a} \perp \mathsf{b}$
5. ***symmetry****:* $MS(\mathsf{a}, \mathsf{b}) = MS(\mathsf{b}, \mathsf{a})$

Proof. For 1, by strict concavity, $S\left(\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}\right) \geq \frac{1}{2}S(\mathsf{a}) + \frac{1}{2}S(\mathsf{b})$, which means $S\left(\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}\right) - \left(\frac{1}{2}S(\mathsf{a}) + \frac{1}{2}S(\mathsf{b})\right) = MS(\mathsf{a}, \mathsf{b}) \geq 0$.

For 2, the concavity is strict and therefore the equality holds if and only if $\mathsf{a} = \mathsf{b}$.

For 3, by the upper variability bound, $S\left(\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}\right) \leq I(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}S(\mathsf{a}) + \frac{1}{2}S(\mathsf{b})$, which means $S\left(\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}\right) - \left(\frac{1}{2}S(\mathsf{a}) + \frac{1}{2}S(\mathsf{b})\right) = MS(\mathsf{a}, \mathsf{b}) \leq I(\frac{1}{2}, \frac{1}{2}) = 1$.

For 4, the upper variability bound is saturated if and only if $\mathsf{a} \perp \mathsf{b}$.

For 5, by commutativity of mixing and of addition the definition of the mixing entropy is symmetric. □

Proposition 1.117. *In discrete and continuous classical cases, the mixing entropy coincides with the Jensen-Shannon divergence. In quantum spaces it coincindes with the quantum Jensen-Shannon divergence.*

Proof. Looking at the definitions of the Jensen-Shannon divergence, in the classical case we have

$$JSD(\mathsf{a}, \mathsf{b}) = S\left(\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}\right) - \frac{1}{2}\left(S(\mathsf{a}) + S(\mathsf{b})\right) = MS(\mathsf{a}, \mathsf{b}),$$

and, similarly, in the quantum case we have

$$QJSD(\mathsf{a}, \mathsf{b}) = S\left(\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}\right) - \frac{1}{2}\left(S(\mathsf{a}) + S(\mathsf{b})\right) = MS(\mathsf{a}, \mathsf{b}).$$

□

Remark. The mixing entropy fails to be a distance function as it does not satisfy the triangle inequality. In the classical and quantum case, in fact, the JSD and QJSD are

the square of a distance function. Given the current axiom, it is unlikely that this result can be generalized to the mixing entropy itself. The issue is that we do not have enough information about the relative entropy of 3 points, apart from the orthogonal case. It is very likely that, on physics grounds, there should be constraints on the mixing entropy between three points.

Though the mixing entropy is not a distance function, we can still show that, like a distance, it decreases as one ensemble approaches another. The notion of two ensembles approaching is given by the convex structure: in a convex space, a mixture of a and b is closer to a than b is to a.

Proposition 1.118. *Let* $\mathsf{a}, \mathsf{b} \in \mathcal{E}$. *Then* $MS(\mathsf{a}, \mathsf{b}) \geq MS(\mathsf{a}, p\mathsf{a}+\bar{p}\mathsf{b})$ *with the equality holding if and only if* $p = 0$ *or* $\mathsf{a} = \mathsf{b}$.

Proof. First let us show that $\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}$ can be expressed as a mixture of $\frac{1}{2}\mathsf{a} + \frac{1}{2}(p\mathsf{a} + \bar{p}\mathsf{b})$ and $p\mathsf{a} + \bar{p}\mathsf{b}$.

$$\begin{aligned}\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b} &= \frac{1}{2}\frac{1-2p+p-2p^2+2p^2}{1-p}\mathsf{a} + \frac{1}{2}(1-2p+2p)\,\mathsf{b} \\ &= \left(\frac{1-2p}{1-p}\frac{1+p}{2} + \frac{2}{2}\frac{p^2}{1-p}\right)\mathsf{a} + \left(\frac{1-2p}{1-p}\frac{\bar{p}}{2} + \frac{2}{2}\frac{p\bar{p}}{1-p}\right)\mathsf{b} \\ &= \frac{1-2p}{1-p}\left[\frac{1}{2}\mathsf{a} + \frac{1}{2}(p\mathsf{a} + \bar{p}\mathsf{b})\right] + \frac{p}{1-p}\left[p\mathsf{a} + \bar{p}\mathsf{b}\right]\end{aligned} \tag{1.119}$$

From strict concavity we have

$$S\left(\frac{1}{2}\mathsf{a} + \frac{1}{2}\mathsf{b}\right) \geq \frac{1-2p}{1-p}S\left(\frac{1}{2}\mathsf{a} + \frac{1}{2}(p\mathsf{a} + \bar{p}\mathsf{b})\right) + \frac{p}{1-p}S(p\mathsf{a} + \bar{p}\mathsf{b}). \tag{1.120}$$

Note that the equality holds only if $p = 0$, in which case the second mixing coefficient is zero.

Now let us show that $p\mathsf{a} + \bar{p}\mathsf{b}$ can be expressed as a mixture of $\frac{1}{2}\mathsf{a} + \frac{1}{2}(p\mathsf{a} + \bar{p}\mathsf{b})$ and b.

$$\begin{aligned}p\mathsf{a} + \bar{p}\mathsf{b} &= p\mathsf{a} + \bar{p}\frac{p+1}{1+p}\mathsf{b} \\ &= \frac{2p}{1+p}\frac{1+p}{2}\mathsf{a} + \left(\frac{2p}{1+p}\frac{\bar{p}}{2} + \frac{\bar{p}}{1+p}\right)\mathsf{b} \\ &= \frac{2p}{1+p}\left[\frac{1}{2}\mathsf{a} + \frac{1}{2}(p\mathsf{a} + \bar{p}\mathsf{b})\right] + \frac{1-p}{1+p}\mathsf{b}\end{aligned} \tag{1.121}$$

From strict concavity we have

$$\begin{aligned} S(p\mathsf{a}+\bar{p}\mathsf{b}) &\geq \frac{2p}{1+p}S\left(\frac{1}{2}\mathsf{a}+\frac{1}{2}(p\mathsf{a}+\bar{p}\mathsf{b})\right)+\frac{1-p}{1+p}S(\mathsf{b}) \\ -S(\mathsf{b}) &\geq \frac{1+p}{1-p}\frac{2p}{1+p}S\left(\frac{1}{2}\mathsf{a}+\frac{1}{2}(p\mathsf{a}+\bar{p}\mathsf{b})\right)-\frac{1+p}{1-p}S(p\mathsf{a}+\bar{p}\mathsf{b}) \\ &= \frac{2p}{1-p}S\left(\frac{1}{2}\mathsf{a}+\frac{1}{2}(p\mathsf{a}+\bar{p}\mathsf{b})\right)-\frac{1+p}{1-p}S(p\mathsf{a}+\bar{p}\mathsf{b}). \end{aligned} \tag{1.122}$$

Note that the equality holds only if $p=0$, in which case the first mixing coefficient is zero.

Putting it all together

$$\begin{aligned} MS(\mathsf{a},\mathsf{b}) &= S\left(\frac{1}{2}\mathsf{a}+\frac{1}{2}\mathsf{b}\right)-\frac{1}{2}S(\mathsf{a})-\frac{1}{2}S(\mathsf{b}) \\ &\geq \frac{1-2p}{1-p}S\left(\frac{1}{2}\mathsf{a}+\frac{1}{2}(p\mathsf{a}+\bar{p}\mathsf{b})\right)+\frac{p}{1-p}S(p\mathsf{a}+\bar{p}\mathsf{b})-\frac{1}{2}S(\mathsf{a}) \\ &+\frac{1}{2}\left[\frac{2p}{1-p}S\left(\frac{1}{2}\mathsf{a}+\frac{1}{2}(p\mathsf{a}+\bar{p}\mathsf{b})\right)-\frac{1+p}{1-p}S(p\mathsf{a}+\bar{p}\mathsf{b})\right] \\ &= \left(\frac{1-2p+p}{1-p}\right)S\left(\frac{1}{2}\mathsf{a}+\frac{1}{2}(p\mathsf{a}+\bar{p}\mathsf{b})\right)-\frac{1}{2}S(\mathsf{a})+\frac{1}{2}\left(\frac{2p-1-p}{1-p}\right)S(p\mathsf{a}+\bar{p}\mathsf{b}) \\ &= S\left(\frac{1}{2}\mathsf{a}+\frac{1}{2}(p\mathsf{a}+\bar{p}\mathsf{b})\right)-\frac{1}{2}S(\mathsf{a})-\frac{1}{2}S(p\mathsf{a}+\bar{p}\mathsf{b}) \\ &= MS(\mathsf{a},p\mathsf{a}+\bar{p}\mathsf{b}). \end{aligned} \tag{1.123}$$

The equality holds only if $p=0$. □

The fact that the mixing entropy decreases from all directions as we get closer to an ensemble allows us to create a notion of open ball, which allows to prove the topology is at least Hausdorff.

Definition 1.124. *Given $\mathsf{a}\in\mathcal{E}$ and $r\in(0,1]$, an entropic open ball is the set of all ensembles for which the mixing entropy from a is within r. That is, $B_r(\mathsf{a})=\{\mathsf{e}\in\mathcal{E}\,|\,MS(\mathsf{a},\mathsf{e})<r\}$.*

Corollary 1.125. *Every entropic open ball is an open set.*

Proof. Since both the entropy and the mixing operation are continuous, the mixing entropy is also continuous. An entropic open ball is the reverse image of an open set through a continuous function and it is therefore an open set. □

Proposition 1.126. *Ensemble spaces are Hausdorff topological spaces.*

Proof. We will prove this in two ways. First, since the topology is second countable, the space is Hausdorff if and only if limits of sequences are unique. Let a_i be a sequence such that $\mathsf{a}_i\to\mathsf{a}$ and $\mathsf{a}_i\to\mathsf{b}$. We have $MS(\mathsf{a}_i,\mathsf{a}_i)=0\to 0$. Since MS is continuous, we also have $MS(\mathsf{a}_i,\mathsf{a}_i)\to MS(\mathsf{a},\mathsf{b})$ which means $MS(\mathsf{a},\mathsf{b})=0$ and therefore $\mathsf{a}=\mathsf{b}$.

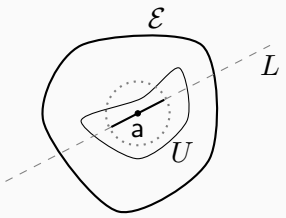
As another way to show it, a space is Hausdorff if and only if any singleton $\{\mathsf{e}\}$ is the intersection of all closed neighborhoods of e. For any e, the intersection of all closed

neighborhoods will contain e, so we just need to show that no other elements can be in all closed neighborhoods. Consider an entropic closed ball centered around e with radius $r > 0$. It will contain an entropic open ball of the same radius and is therefore a closed neighborhood of e. Given an element $\mathsf{a} \neq \mathsf{e}$, a closed ball with radius less than $MS(\mathsf{e},\mathsf{a})$ will not contain a. Therefore, the only element contained in all entropic closed balls is e, which means the intersection of all closed neighborhoods of e contains only e. □

Remark. Intuitively, we would expect the entropic balls to be convex. This is true in both classical and quantum cases as the JSD is convex. However, the current axioms allow us to use a different entropy function in a space that is not orthogonally decomposable: the extreme points are not orthogonal. In this case, one is able to create open balls that are not convex.

Whether this is physically relevant or not depends on whether there are additional physical constraints on the mixing entropy between three points, which is very likely.

Proposition 1.127. *Let $U \subseteq \mathcal{E}$ be an open set. Let $\mathsf{a} \in U$ and $L \subseteq \mathcal{E}$ be a line containing a. Then there is an $r \in (0,1]$ such that for all $\mathsf{e} \in L$ such that $MS(\mathsf{a},\mathsf{e}) < r$ we have $\mathsf{e} \in U$.*



Proof. Let $U \subseteq \mathcal{E}$ be an open set, $\mathsf{a} \in U$ and $L \subseteq \mathcal{E}$ be a line containing a. Take $\mathsf{b} \in L$ such that $\mathsf{b} \neq \mathsf{a}$. Let $f : [0,1] \to \mathcal{E}$ be such that $f(p) \mapsto p\mathsf{a} + \bar{p}\mathsf{b}$. Since the mixing operation is continuous, then f is also continuous. This means that $f^{-1}(U)$ is an open subset of $[0,1]$. Since $\mathsf{a} \in U$, $1 \in f^{-1}(U)$ and therefore there will be an open interval $(\lambda, 1]$ such that $(\lambda, 1] \subseteq f^{-1}(U)$. Let $\mathsf{c} = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b}$ and $r = MS(\mathsf{a},\mathsf{c})$. By 1.118, we have that, for all $p \in (0,1]$, $MS(\mathsf{a}, p\mathsf{a} + \bar{p}\mathsf{c}) < r$. We also have $p\mathsf{a} + \bar{p}\mathsf{c} = p\mathsf{a} + \bar{p}(\lambda\mathsf{a} + \bar{\lambda}\mathsf{b}) = (p + \bar{p}\lambda)\mathsf{a} + \bar{p}\bar{\lambda}\mathsf{b}$ where $p + \bar{p}\lambda \in (\lambda, 1]$. Therefore $p\mathsf{a} + \bar{p}\mathsf{c} \in f((\lambda, 1]) \subseteq U$.

If a is an extreme point of L (i.e. there is no $\mathsf{d} \in L$ such that $p\mathsf{b} + \bar{p}\mathsf{d} = \mathsf{a}$), then we are done. If a is not an extreme point, then we can find $\mathsf{d} \in L$ such that $p\mathsf{b} + \bar{p}\mathsf{d} = \mathsf{a}$. Repeat the previous procedure with d to find another r' and let $\hat{r}$ be the minimum between the two. Then we have that $\hat{r} \in (0,1]$ such that $\mathsf{e} \in U$ for all $\mathsf{e} \in L$ such that $MS(\mathsf{a},\mathsf{e}) < \hat{r}$. □

There are a number of conjectures on the topology generated by the entropic open ball that are probably all related. They are summarized here with some notes on what has been understood so far.

Conjecture 1.128. *For every sequence $\mathsf{a}_i \in \mathcal{E}$, $\mathsf{a}_i \to \mathsf{a}$ if and only if $MS(\mathsf{a},\mathsf{a}_i) \to 0$.*

Conjecture 1.129. *The topology of the ensemble space is generated by the entropic open balls.*

Conjecture 1.130. *Let $U \subseteq \mathcal{E}$ be an open set and $\mathsf{a} \in U$. Then there exists an $r \in (0,1]$ such that $B_r(\mathsf{a}) \subseteq U$.*

Remark. These conjectures are likely the same. If the topology is generated by the entropic open balls, then convergence in mixing entropy is the convergence criterion associated with the topology. It also means that we can generate every open set with entropic open balls by fitting them inside every open set.

Note that all entropic open balls are open sets in the topology, since the mixing entropy is a continuous function. It suffices to prove that all open sets can be generated by open balls. We have shown that given a point in an open set we can find a finite range of mixing entropy along each direction. However, since we do not know if there is a non-zero infimum across all directions, we do not know whether we can "fit" an entropic open ball.

Similarly, it is clear that if a sequence converges, the mixing entropy goes to zero. However, we do not know whether there is an open set for which the mixing entropy goes to zero but the sequence does not converge. This would mean that a sequence always eventually enters an entropic open ball of any size, but it may not enter all open sets.

Conjecture 1.131. *Let* $\mathsf{a}, \mathsf{b}_i \in \mathcal{E}$ *and* $\mathsf{c}_i = p\mathsf{a} + \bar{p}\mathsf{b}_i$ *for some* $p \in (0,1)$*. Then* $\mathsf{b}_i \to \mathsf{b}$ *if and only if* $\mathsf{c}_i \to \mathsf{c}$*.*

Conjecture 1.132 (Affine combinations are continuous)**.** *Let* $\mathsf{e} \in \mathcal{E}$ *and let* $p \in [0,1]$*. Let* $+_{p\mathsf{a}} : \mathcal{E} \to \mathcal{E}$ *be the curried function* $+_{p\mathsf{a}}(\mathsf{b}) = p\mathsf{a} + \bar{p}\mathsf{b}$*. The function is a homeomorphism between* $\mathcal{E}$ *and* $+_{p\mathsf{a}}(\mathcal{E})$*.*

Conjecture 1.133. *Mixing is an open map.*

Conjecture 1.134. *An ensemble space embeds continuously into a topological vector space.*

Conjecture 1.135. *Let* $+_{\mathsf{ab}} : [0,1] \to \mathcal{E}$ *be the curried function* $+_{\mathsf{ab}}(p) = p\mathsf{a} + \bar{p}\mathsf{b}$*. The function is a homeomorphism between* $[0,1]$ *and* $+_{\mathsf{ab}}([0,1])$*.*

Remark. These are all about showing that mixing allows a continuous inverse. It is easy to show that mixing gives us a continuous bijection. The problem is showing that the inverse is continuous.

Continuity of mixtures only shows that we can stretch open sets to bigger open sets. What we would need is to show that we can shrink open sets to open sets. For example, suppose U is an open neighborhood of a. The set $V = p\mathsf{a} + \bar{p}U$ is a set that contains a and it is a subset of U. The question is whether a sequence can stay in U and converge to a without ever entering V.

Another open question is whether the mixing entropy, or the entropy directly, can be related to a generalized notion of inner product.

Entropic metric

Following the first observation that the mixing entropy is a pseudo-distance, the second observation is that the strict concavity of the entropy forces the Hessian to be negative definite. The negation of the Hessian, then, is a positive definite function of two variations. In general, the Hessian of a scalar function is not a tensor. However, the ensemble space is a linear space and that linearity has physical significance. It is only for coordinates linear with respect to the mixing coefficient that the convex combination of coordinates will equal the statistical mixture of the corresponding ensembles. Therefore we can define a metric tensor which corresponds to the negative Hessian calculated on that linear structure. When using non-linear coordinates,

as long as the coordinates are smooth, one can always make the appropriate transformation to linear coordinates.

Remark. In this section we will assume that an ensemble space embeds continuously into a topological vector space, even though we have not yet proved it. We are going to talk about variations $\delta\mathsf{e}$ on the space, even though it is not yet clear exactly which mathematical approach is best to make variations well-defined. We are also going to talk about a metric tensor even though the space is not, in general, a manifold. All these issues will need to be resolved in due time.

Definition 1.136. *An ensemble space is **smooth** if the entropy is twice differentiable with respect to the mixing coefficients.*

Proposition 1.137. *Discrete and continuous classical ensemble spaces and quantum ensemble spaces are smooth.*

Proof. In both the discrete classical case and the quantum case the entropy of an ensemble is the Shannon entropy of a decomposition in terms of pure states. The Shannon entropy is a smooth function of the coefficients. For the continuous classical case, the differential entropy is a smooth function of the probability density, which is linear under mixture. Therefore the entropy is always smooth with respect to mixing coefficients. □

Definition 1.138. *Assuming $\mathcal{E}$ embeds in a topological vector space and given $\mathsf{e} \in \mathcal{E}$, a **variation** $\delta\mathsf{e}$ of e is a vector in the ambient space such that $\forall t \in [0,1]\ \mathsf{e} + t\delta\mathsf{e} \in \mathcal{E}$. The space of all variations at e is noted T_{e}. Unless otherwise noted, we assume the variation is expressed on the affine structure. Note that it can be re-expressed through a non-linear map as long as the map is differentiable (i.e. it maps variations to variations).*

Remark. Note that since an ensemble space is a convex subset of a real vector space, coordinate systems that are linear with respect to the vector space are privileged. It is only in these coordinates, in fact, that linear combinations correspond to mixtures. The strict concavity of the entropy is therefore guaranteed in these coordinates and only these coordinates. Given the special physical significance of these coordinates, all differential objects and properties will be defined in these coordinates.

Definition 1.139. *Given an ensemble $\mathsf{e} \in \mathcal{E}$ and a variation $\delta\mathsf{e}$ defined at that point, the **norm** of $\delta\mathsf{e}$ is given by*

$$\|\delta\mathsf{e}\|_{\mathsf{e}} = \sqrt{8MS(\mathsf{e}, \mathsf{e} + \delta\mathsf{e})}.$$

*The **metric tensor** (i.e. the inner product between $\delta\mathsf{e}_1, \delta\mathsf{e}_2 \in T_{\mathsf{e}}$) is given by*

$$g_{\mathsf{e}}(\delta\mathsf{e}_1, \delta\mathsf{e}_2) = \frac{1}{2}\left(\|\delta\mathsf{e}_1 + \delta\mathsf{e}_2\|_{\mathsf{e}}^2 - \|\delta\mathsf{e}_1\|_{\mathsf{e}}^2 - \|\delta\mathsf{e}_2\|_{\mathsf{e}}^2\right).$$

Theorem 1.140. *Let $\mathcal{E}$ be a smooth ensemble space. Then, on the affine structure, we have*

$$\|\delta\mathsf{e}\|_{\mathsf{e}}^2 = -\frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}, \delta\mathsf{e})$$

and

$$g_{\mathsf{e}}(\delta\mathsf{e}_1, \delta\mathsf{e}_2) = -\frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_1, \delta\mathsf{e}_2).$$

Proof. To recover the first two expressions, we simply have to calculate the leading term. Since the entropy is twice differentiable, on the affine structure, we can expand it as

$$S(\mathsf{e} + \delta\mathsf{e}) = S(\mathsf{e}) + \frac{\partial S}{\partial \mathsf{e}}\delta\mathsf{e} + \frac{1}{2}\frac{\partial^2 S}{\partial \mathsf{e}^2}\delta\mathsf{e}\delta\mathsf{e} + O(\delta\mathsf{e}^3). \tag{1.141}$$

Expanding the definition of MS, we have

$$\begin{aligned}
MS(\mathsf{e}, \mathsf{e} + \delta\mathsf{e}) &= S\left(\frac{1}{2}\mathsf{e} + \frac{1}{2}(\mathsf{e} + \delta\mathsf{e})\right) - \frac{1}{2}S(\mathsf{e}) - \frac{1}{2}S(\mathsf{e} + \delta\mathsf{e}) \\
&= S\left(\mathsf{e} + \frac{1}{2}\delta\mathsf{e}\right) - \frac{1}{2}S(\mathsf{e}) - \frac{1}{2}S(\mathsf{e} + \delta\mathsf{e}) \\
&= S(\mathsf{e}) + \frac{\partial S}{\partial \mathsf{e}}\frac{1}{2}\delta\mathsf{e} + \frac{1}{2}\frac{\partial^2 S}{\partial \mathsf{e}^2}\frac{1}{2}\delta\mathsf{e}\frac{1}{2}\delta\mathsf{e} + O(\delta\mathsf{e}^3) \\
&\quad - \frac{1}{2}S(\mathsf{e}) - \frac{1}{2}\left(S(\mathsf{e}) + \frac{\partial S}{\partial \mathsf{e}}\delta\mathsf{e} + \frac{1}{2}\frac{\partial^2 S}{\partial \mathsf{e}^2}\delta\mathsf{e}\delta\mathsf{e} + O(\delta\mathsf{e}^3)\right) \\
&= S(\mathsf{e}) + \frac{1}{2}\frac{\partial S}{\partial \mathsf{e}}\delta\mathsf{e} + \frac{1}{8}\frac{\partial^2 S}{\partial \mathsf{e}^2}\delta\mathsf{e}\delta\mathsf{e} \\
&\quad - S(\mathsf{e}) - \frac{1}{2}\frac{\partial S}{\partial \mathsf{e}}\delta\mathsf{e} - \frac{1}{4}\frac{\partial^2 S}{\partial \mathsf{e}^2}\delta\mathsf{e}\delta\mathsf{e} + O(\delta\mathsf{e}^3) \\
&= -\frac{1}{8}\frac{\partial^2 S}{\partial \mathsf{e}^2}\delta\mathsf{e}\delta\mathsf{e} + O(\delta\mathsf{e}^3).
\end{aligned} \tag{1.142}$$

Therefore

$$\|\delta\mathsf{e}\|^2 = 8MS(\mathsf{e}, \mathsf{e} + \delta\mathsf{e}) = -\frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}, \delta\mathsf{e}).$$

We can now substitute the norm in the definition of the metric tensor. We have

$$\begin{aligned}
g_{\mathsf{e}}(\delta\mathsf{e}_1, \delta\mathsf{e}_2) &= \frac{1}{2}\left(\|\delta\mathsf{e}_1 + \delta\mathsf{e}_2\|^2 - \|\delta\mathsf{e}_1\|^2 - \|\delta\mathsf{e}_2\|^2\right) \\
&= \frac{1}{2}\left(-\frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_1 + \delta\mathsf{e}_2, \delta\mathsf{e}_1 + \delta\mathsf{e}_2) + \frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_1, \delta\mathsf{e}_1) + \frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_2, \delta\mathsf{e}_2)\right) \\
&= -\frac{1}{2}\left(\frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_1, \delta\mathsf{e}_1) + \frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_1, \delta\mathsf{e}_2) + \frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_2, \delta\mathsf{e}_1) + \frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_2, \delta\mathsf{e}_2)\right. \\
&\quad \left. - \frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_1, \delta\mathsf{e}_1) - \frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_2, \delta\mathsf{e}_2)\right) \\
&= -\frac{\partial^2 S}{\partial \mathsf{e}^2}(\delta\mathsf{e}_1, \delta\mathsf{e}_2).
\end{aligned} \tag{1.143}$$

□

We now show that the metric tensor we defined reduces to the Fisher-Rao metric in the classical case and to its quantum equivalent in the quantum case.

Proposition 1.144. *For a continuous classical ensemble space, the metric corresponds to the Fisher-Rao metric.*

Proof. Let $\mathcal{E}$ be a continuous classical ensemble space. Each ensemble is a classical probability density ρ. Let $V \subseteq \mathcal{E}$ be a manifold of probability distributions over X parametrized by θ^i. The Fisher-Rao metric is defined as:

$$g_{ij} = -\int_X \frac{\partial^2 \log \rho}{\partial\theta^i \partial\theta^j} \rho dx. \tag{1.145}$$

where log will be the natural logarithm throughout this calculation. Recall that for a continuous classical ensemble, the entropy is given by

$$S(\rho) = -\int_X \rho \log \rho dx. \tag{1.146}$$

Let us now calculate the first two terms in the Taylor expansion around ρ with a variation $\delta\rho$. Recall that

$$\begin{aligned} \log(x+dx) &= \log(x) + d_x \log x\, dx + \frac{1}{2} d_x d_x \log x\, dx^2 + O(dx^3) \\ &= \log(x) + \frac{1}{x} dx - \frac{1}{2}\frac{1}{x^2} dx^2 + O(dx^3). \end{aligned} \tag{1.147}$$

We have

$$\begin{aligned} S(\rho+\delta\rho) &= -\int_X (\rho+\delta\rho)\log(\rho+\delta\rho) dx \\ &= -\int_X (\rho+\delta\rho)\left[\log\rho + \frac{1}{\rho}\delta\rho - \frac{1}{2\rho^2}\delta\rho^2 + O(\delta\rho^3)\right] dx \\ &= -\int_X \rho\log\rho dx - \int_X [\log\rho + 1]\, dx\delta\rho - \int_X \left[\frac{1}{\rho} - \frac{\rho}{2\rho^2}\right] dx\delta\rho^2 + \int_X dx O(\delta\rho^3) \\ &= -\int_X \rho\log\rho dx - \int_X [\log\rho + 1]\, dx\delta\rho - \frac{1}{2}\int_X \frac{1}{\rho} dx\delta\rho^2 + \int_X dx O(\delta\rho^3) \end{aligned} \tag{1.148}$$

$$\begin{aligned} \frac{\partial^2 S}{\partial\rho^2}(\delta\rho,\delta\rho) &= -\int_X \frac{1}{\rho}\delta\rho^2 dx = -\int_X \frac{1}{\rho}\delta\rho^2 dx + 0 = -\int_X \frac{1}{\rho}\delta\rho^2 dx + \delta^2(1) \\ &= -\int_X \frac{1}{\rho}\delta\rho^2 dx + \delta^2 \int_X \rho dx = -\int_X \frac{1}{\rho}\delta\rho^2 dx + \int_X \delta^2 \rho dx \\ &= \int_X \rho dx \left[-\frac{1}{\rho^2}\delta\rho^2 + \frac{1}{\rho}\delta^2\rho\right] = \int_X \rho dx\, \delta\left[\frac{1}{\rho}\delta\rho\right] \\ &= \int_X \rho dx\, \delta^2 \log\rho. \end{aligned} \tag{1.149}$$

Let us consider a family of ensembles charted by a set of parameters θ^i, not necessarily forming a linear chart. We have

$$g_{\mathsf{e}}(d\theta^i, d\theta^j) = -\frac{\partial^2 S}{\partial \rho^2}\left(\frac{\partial \rho}{\partial \theta^i}d\theta^i, \frac{\partial \rho}{\partial \theta^j}d\theta^j\right) = -\int_X \rho dx \frac{\partial^2 \log \rho}{\partial \theta^i \partial \theta^j} d\theta^i d\theta^j \tag{1.150}$$

which recovers the Fisher-Rao metric.

This same calculation can be done in the case of a discrete classical space. □

Remark. For quantum mechanics, the situation is more complicated as there are different definitions of Fisher metrics[a]. Additionally, we will take steps in the calculations that are already present in established literature, though the mathematical conditions for those steps to be well-defined are unclear. We will therefore just show a general connection, without worrying about the mathematical details.

Proposition 1.151. *For a quantum ensemble space, the metric corresponds to the Bures metric and the quantum Fisher information metric.*

Proof. Since we are working in a quantum ensemble space, each ensemble is a density operator ρ. The entropy is given by the von Neumann entropy

$$S(\rho) = -\operatorname{tr}(\rho \log \rho). \tag{1.152}$$

where log will be the natural logarithm throughout this calculation. We take the first variation and have

$$\begin{aligned} \delta S(\rho) &= -\delta \operatorname{tr}(\rho \log \rho) = -\operatorname{tr}(\delta\rho \log \rho + \rho \delta \log \rho) \\ &= -\operatorname{tr}\left(\delta\rho \log \rho + \rho\rho^{-1}\delta\rho\right) \\ &= -\operatorname{tr}\left((\log \rho + 1)\,\delta\rho\right). \end{aligned} \tag{1.153}$$

We take the second variation and have

$$\begin{aligned} \delta^2 S(\rho) &= -\delta \operatorname{tr}\left((\log \rho + 1)\,\delta\rho\right) \\ &= -\operatorname{tr}\left(\delta\left(\log \rho + 1\right)\delta\rho + \left(\log \rho + 1\right)\delta\delta\rho\right) \\ &= -\operatorname{tr}\left(\rho^{-1}\delta\rho\delta\rho + \left(\log \rho + 1\right)\delta\delta\rho\right). \end{aligned} \tag{1.154}$$

Note that $\delta\rho$ is defined in a linear chart, therefore $\delta\delta\rho = 0$. Therefore we have

$$\frac{\partial^2 S}{\partial \rho^2}(\delta\rho, \delta\rho) = -\operatorname{tr}\left(\rho^{-1}\delta\rho\delta\rho\right). \tag{1.155}$$

Let us consider a family of ensembles charted by a set of parameters θ^i, not necessarily forming a linear chart. We have

$$g_{\mathsf{e}}(d\theta^i, d\theta^j) = -\frac{\partial^2 S}{\partial \rho^2}\left(\frac{\partial \rho}{\partial \theta^i}d\theta^i, \frac{\partial \rho}{\partial \theta^j}d\theta^j\right) = \operatorname{tr}\left(\rho^{-1}\frac{\partial \rho}{\partial \theta^i}\frac{\partial \rho}{\partial \theta^j}\right) d\theta^i d\theta^j \tag{1.156}$$

which recovers one version of the quantum Fisher-Rao metric.

Noting that the right logarithmic derivative (RLD) L^R satisfies $\frac{\partial \rho}{\partial \theta} = \rho L^R$, we can write

$$g_{\mathsf{e}}(d\theta^i, d\theta^j) = \mathrm{tr}\left(\rho^{-1}\frac{\partial \rho}{\partial \theta^i}\frac{\partial \rho}{\partial \theta^j}\right) d\theta^i d\theta^j = \mathrm{tr}(\rho^{-1}\rho L_i^R \rho L_j^R) d\theta^i d\theta^j = \mathrm{tr}(\rho L_i^R L_j^R) d\theta^i d\theta^j \quad (1.157)$$

which recovers the RLD Fisher information. □

Remark. The feedback we received on the above derivation is contradictory, as some say we are handling the possible non-commutativity between ρ and its variation $\delta\rho$ correctly and others say we aren't. Note that in all steps we are using the cyclical nature of the trace, and not commuting ρ and its variation $\delta\rho$.

Remark. Note that these results relate also to the Ruppeiner metric, which is the negation of the Hessian of the entropy. They also relate to the Hessian metric as described in Shima, H. (2013). Lecture Notes in Computer Science, vol 8085.

[a] For details, see Vishal Katariya and Mark M Wilde 2021 New J. Phys. 23 073040 or Watanabe, Y. (2014). Quantum Estimation Theory.

1.7 State capacity

In this section we are going to develop a generalized version of the count of states. In classical statistical mechanics, the entropy for a uniform distribution ρ_U over a region of phase space U is given by $S(\rho_U) = \log \mu(U)$, where μ is the count of states given by the Liouville measure. Since entropy is our starting point, we will define the state capacity of a set of ensembles as the highest exponential of the entropy achievable by mixing those ensembles. This definition works in general, and recovers both the classical Liouville measure and the dimensionality of a quantum subspace.

Given an ensemble, we want to quantify the number of distinguishable cases over which the ensemble is spread. Since entropy characterizes the variability of the preparations of the ensemble, greater variability means the ensemble is spread over more cases. Therefore the count of distinguishable states must be a monotonic function of the entropy. The question is what function.

The first hint that the exponential is the right function is the relationship $S(\rho_U) = \log \mu(U)$ in classical statistical mechanics. Another hint is that we will want this size to be multiplicative over independent distributions. That is, if we have two ensembles $\rho_1 \in \mathcal{E}_1$ and $\rho_2 \in \mathcal{E}_2$ of two different ensemble spaces, we can imagine the composite system $\rho = \rho_1\rho_2$ representing the ensemble of the two independent ensembles. While the entropy should be additive, that is $S(\rho) = S_1(\rho_1) + S_2(\rho_2)$, the count of states should be multiplicative, that is $\mu(\rho) = \mu_1(\rho_1)\mu_2(\rho_2)$. This suggests the relationship $S(\rho) = \log \mu(\rho)$ for all ensembles. The other hint is that the spread can be, at most, additive. If we mix two ensembles, the spread can be at most over the configurations of both and not more. The following shows that the exponential of the entropy has exactly that property.[12]

[12] Note that the state capacity for a single element corresponds to the "ensemble volume" defined in M.J.W. Hall 1999, Phys. Rev. A 59, 2602 and more recently M.J.W. Hall 2018 J. Phys. A: Math. Theor. 51 364001. It is connected to various uncertainty relationships, including a stronger form of the quantum uncertainty principle. It would be interesting to understand what results can be generalized to the ensemble space.

Proposition 1.158 (Exponential entropy subadditivity)**.** *Let* $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ *and* $\mathsf{e} = p\mathsf{a} + \bar{p}\mathsf{b}$ *for some* $p \in [0,1]$*. Then* $2^{S(\mathsf{e})} \leq 2^{S(\mathsf{a})} + 2^{S(\mathsf{b})}$*, with the equality holding if and only if* $\mathsf{a} \perp \mathsf{b}$ *and* $p = \frac{2^{S(\mathsf{a})}}{2^{S(\mathsf{a})}+2^{S(\mathsf{b})}}$*.*

Proof. If p is fixed, the upper variability bound of entropy is saturated only if a and b are orthogonal by definition. Therefore the entropy maximum for the mixed ensemble will be achieved when the elements are orthogonal, for some value of p.

Now fix the entropy of a and b to some values $S_a = S(\mathsf{a})$ and $S_b = S(\mathsf{b})$. The entropy of the mixture depends only on p, so we need to find the p that maximizes the expression. Since a and b are orthogonal, $S(p\mathsf{a} + \bar{p}\mathsf{b}) = -p\log p - \bar{p}\log\bar{p} + pS_a + \bar{p}S_b$ which is a smooth function of p. Let us find the maximum, which is a stationary point.

$$\begin{aligned}
0 &= \frac{d}{dp}S(p\mathsf{a} + \bar{p}\mathsf{b}) = \frac{d}{dp}\left(-p\log p - \bar{p}\log\bar{p} + pS_a + \bar{p}S_b\right) \\
&= -\log p - 1 + \log\bar{p} + 1 + S_a - S_b \\
\log\frac{p}{\bar{p}} &= \log 2^{S_a} - \log 2^{S_b} \\
\log\frac{p}{1-p} &= \log\frac{2^{S_a}}{2^{S_b}} \\
p2^{S_b} &= (1-p)2^{S_a} \\
p(2^{S_a} + 2^{S_b}) &= 2^{S_a} \\
p &= \frac{2^{S_a}}{2^{S_a} + 2^{S_b}}
\end{aligned} \tag{1.159}$$

Note that since the entropy is strictly concave, the stationary point must correspond to a maximum. Having found the value of p that maximizes the entropy, we can calculate the maximum entropy.

$$\begin{aligned}
\bar{p} &= 1 - \frac{2^{S_a}}{2^{S_a} + 2^{S_b}} = \frac{2^{S_b}}{2^{S_a} + 2^{S_b}} \\
S(p\mathsf{a} + \bar{p}\mathsf{b}) &= -p\log p - \bar{p}\log\bar{p} + pS_a + \bar{p}S_b \\
&= -\frac{2^{S_a}}{2^{S_a} + 2^{S_b}}\log\frac{2^{S_a}}{2^{S_a} + 2^{S_b}} - \frac{2^{S_b}}{2^{S_a} + 2^{S_b}}\log\frac{2^{S_b}}{2^{S_a} + 2^{S_b}} \\
&+ \frac{2^{S_a}}{2^{S_a} + 2^{S_b}}\log 2^{S_a} + \frac{2^{S_b}}{2^{S_a} + 2^{S_b}}\log 2^{S_b} \\
&= \frac{2^{S_a}}{2^{S_a} + 2^{S_b}}\log\left(2^{S_a} + 2^{S_b}\right) + \frac{2^{S_b}}{2^{S_a} + 2^{S_b}}\log\left(2^{S_a} + 2^{S_b}\right) \\
&= \frac{2^{S_a} + 2^{S_b}}{2^{S_a} + 2^{S_b}}\log\left(2^{S_a} + 2^{S_b}\right) \\
\log 2^{S(p\mathsf{a}+\bar{p}\mathsf{b})} &= \log\left(2^{S_a} + 2^{S_b}\right) \\
2^{S(p\mathsf{a}+\bar{p}\mathsf{b})} &= 2^{S_a} + 2^{S_b}
\end{aligned} \tag{1.160}$$

Therefore the maximum entropy obtainable through a mixture is $S(p\mathsf{a} + \bar{p}\mathsf{b}) = \log(2^{S(\mathsf{a})} + 2^{S(\mathsf{b})})$ which is obtained when a and b are orthogonal and $p = \frac{2^{S(\mathsf{a})}}{2^{S(\mathsf{a})}+2^{S(\mathsf{b})}}$. □

Remark. We have proved that the exponential of the entropy gives us subadditivity. Ideally, we should prove that this is the only function of the entropy with that characteristic.

Conjecture 1.161. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function such that $f(S(p\mathsf{a} + \bar{p}\mathsf{b})) \leq f(S(\mathsf{a})) + f(S(\mathsf{b}))$ and the equality can be verified in some condition. Then $f(x) = \kappa^x$ for some arbitrary constant κ.*

Given a set of ensembles, we can ask what spread is reachable by a mixture in terms of the number of distinguishable states. We call this the state capacity as it represents the maximum potential spread reachable by the set, and because it turns out to be a non-additive measure.[13] The state capacity is monotone, subadditive and recovers additivity over orthogonal sets.

Definition 1.162. *Let $A \subseteq \mathcal{E}$ be a subset of an ensemble space. The **state capacity** of A is defined as* $\operatorname{scap}(A) = \sup(2^{S(\operatorname{hull}(A))} \cup \{0\})$.

Corollary 1.163. *The state capacity is a convex supremum of the exponential of the entropy with 0 for the empty set.*

Corollary 1.164. *Given a set $A \subseteq \mathcal{E}$, the state capacity of all hulls is the same. That is,* $\operatorname{scap}(\operatorname{hull}(A)) = \operatorname{scap}(\operatorname{shull}(A)) = \operatorname{scap}(\operatorname{chull}(A))$.

Proof. Since the entropy is continuous, by 1.53 $\operatorname{scap}(A) = \operatorname{scap}(\operatorname{hull}(A)) = \operatorname{scap}(\operatorname{shull}(A)) = \operatorname{scap}(\operatorname{chull}(A))$. □

Proposition 1.165. *The state capacity is a set function that is*

1. *non-negative:* $\operatorname{scap}(A) \in [0, +\infty]$
2. *increasing:* $A \subseteq B \implies \operatorname{scap}(A) \leq \operatorname{scap}(B)$
3. *subadditive:* $\operatorname{scap}(A \cup B) \leq \operatorname{scap}(A) + \operatorname{scap}(B)$
4. *additive over orthogonal sets:* $A \perp B \implies \operatorname{scap}(A \cup B) = \operatorname{scap}(A) + \operatorname{scap}(B)$
5. *continuous from below.*

Proof. 1. By 1.51, $\operatorname{scap}(A) \in [0, \sup(2^{S(\mathcal{E})})] \subseteq [0, +\infty]$.

2. By 1.51.

3. Let $A, B \subseteq \mathcal{E}$ and let $\mathsf{e} = p\mathsf{a} + \bar{p}\mathsf{b}$ for some $p \in [0, 1]$, $\mathsf{a} \in A$ and $\mathsf{b} \in B$. By 1.158 and the definition of state capacity, $2^{S(\mathsf{e})} \leq 2^{S(\mathsf{a})} + 2^{S(\mathsf{b})} \leq \operatorname{scap}(A) + \operatorname{scap}(B)$. Since this is true for any finite mixture of elements of A and B, it will be true for any element of $\operatorname{hull}(A \cup B)$. Consequently, the supremum of the exponential entropy cannot exceed the sum of the state capacities. Therefore $\operatorname{scap}(A \cup B) \leq \operatorname{scap}(A) + \operatorname{scap}(B)$ (i.e. the state capacity is subadditive).

4. Let $A, B \subseteq \mathcal{E}$ be two orthogonal subsets. Let $\{\mathsf{a}_i\} \subseteq A$ be a sequence of ensembles such that $2^{S(\mathsf{a}_i)} \to \operatorname{scap}(A)$ and let $\{\mathsf{b}_i\} \subseteq B$ be a sequence of ensembles such that $2^{S(\mathsf{b}_i)} \to$

[13] In some literature, capacities are non-additive measures.

$\operatorname{scap}(B)$. Consider $\mathsf{e}_i = p_i\mathsf{a}_i + \bar{p}_i\mathsf{b}_i$ where $p_i = \frac{2^{S(\mathsf{a}_i)}}{2^{S(\mathsf{a}_i)}+2^{S(\mathsf{b}_i)}}$. Then, by 1.158, $2^{S(\mathsf{e}_i)} = 2^{S(\mathsf{a}_i)} + 2^{S(\mathsf{b}_i)}$. This means that $2^{S(\mathsf{e}_i)} \to \operatorname{scap}(A)+\operatorname{scap}(B)$. Therefore $\operatorname{scap}(A\cup B)$ must be at least $\operatorname{scap}(A)+\operatorname{scap}(B)$. Combining with the previous result, $\operatorname{scap}(A\cup B) = \operatorname{scap}(A)+\operatorname{scap}(B)$. Therefore the state capacity, as a set function, is additive over orthogonal sets of ensembles.

5. By 1.51. □

Remark. The state capacity is not continuous from above. Let $\mathcal{E}$ be an ensemble space and $\mathsf{a},\mathsf{b} \in \mathcal{E}$ two distinct ensembles. Without loss of generality, let $S(\mathsf{a}) \leq S(\mathsf{b})$. Then $S(p\mathsf{a}+\bar{p}\mathsf{b}) \geq S(\mathsf{a})$ for all $p \in [0,1]$. Let $\{\mathsf{e}_i\} \subset \operatorname{hull}(\{\mathsf{a},\mathsf{b}\})$ be a sequence of mixtures of a and b. Let $A_j = \{\mathsf{e}_i \,|\, i \geq j\}$ be the family of sets that contains all the elements of the sequence starting from j respectively. This is a decreasing sequence, since $A_j \supseteq A_{j+i}$. We have that $\operatorname{scap}(A_j) \geq 2^{S(\mathsf{a})} > 0$ for all j, which means that $\lim_{i\to\infty} \operatorname{scap}(A_i) \geq 2^{S(\mathsf{a})} > 0$. However, $\bigcap A_j = \varnothing$ which means $\operatorname{scap}(\lim_{i\to\infty} A_i) = 0$. This means that, in general, $\operatorname{scap}(\lim_{i\to\infty} A_i) \neq \lim_{i\to\infty} \operatorname{scap}(A_i)$ for a decreasing sequence.

The state capacity is additive over orthogonal sets. Intuitively, orthogonal sets correspond to mutually exclusive events, so it makes sense that the count of states is additive. We would expect the converse to be true as well: if the count of states is additive, the sets correspond to mutually exclusive events and therefore the sets are orthogonal. This leads to the following conjecture:

Conjecture 1.166. *The state capacity is additive only on orthogonal sets. That is, if* $\operatorname{scap}(A) < \infty$, $\operatorname{scap}(B) < \infty$ *and* $\operatorname{scap}(A\cup B) = \operatorname{scap}(A)+\operatorname{scap}(B)$, *then* $A \perp B$.

One should be able to prove that if the state capacity adds, then the states with highest entropy satisfy the upper variability bound and are orthogonal. What needs to be shown is that this is enough to say all elements are orthogonal. There is a series of nice results connected to this to expand the whole theory. If $\mathsf{a} \perp \mathsf{b}$ then a is orthogonal to all the components of b. If a is the ensemble with maximum entropy in $\operatorname{chull}(A)$, it should be true that all elements of the hull are components of a. One should be able to show that, given two sequences with a convergent entropy and whose elements are orthogonal to each other, one can create another sequence that converges to an entropy higher than both of them. Another useful tool would be a type of sequence within $\operatorname{hull}(A)$ whose entropy always increases, tends to the supremum and each element is a component of the following. Also, the state with maximal entropy (if it exists) is an internal point. All internal points are in the σ-hull. We leave these series of conjectures for future work.

We now show that the state capacity recovers a notion of number of cases in all three target spaces. In the discrete classical case it recovers the number of points; in the continuous classical case, it recovers the Liouville volume; in the quantum case, it recovers the number of orthogonal pure states.

Proposition 1.167. *Let* $\mathcal{E}$ *be a discrete classical ensemble space. Let* $U \subseteq \{s_i\}$ *be a subset of the corresponding (extreme) points and* A *be the set of probability distributions whose support is a subset of* U. *Then* $\operatorname{scap}(A) = \#U$.

Proof. Let A be the set of probability distributions defined over a discrete countable set U. The highest entropy is achieved by the uniform distribution, which equals $\log(\#U)$.

Therefore $\text{scap}(A) = 2^{\log(\#U)} = \#U$. □

Proposition 1.168. *Let $\mathcal{E}$ be a continuous classical ensemble space. Let $U \subseteq X$ be a subset of the corresponding symplectic manifold (i.e. phase space) and A be the set of probability distributions whose support is a subset of U. Then* $\text{scap}(A) = \mu(U)$ *where $\mu(U)$ is the Liouville measure.*

Proof. Let A be the set of probability distributions whose support is a subset of U. The highest entropy is achieved by the uniform distribution, which equals $\log(\mu(U))$. Therefore $\text{scap}(A) = 2^{\log(\mu(U))} = \mu(U)$. □

Proposition 1.169. *Let $\mathcal{E}$ be a quantum ensemble space. Let $U \subseteq \mathcal{H}$ be a subspace of the corresponding Hilbert space and A be the set of density operators defined on that subspace (i.e. zero eigenvalues outside). Then* $\text{scap}(A) = \dim(U)$ *where* $\dim(U)$ *is the dimensionality of the subspace.*

Proof. Let A be the set of density operators defined on a finite-dimensional subspace U. The highest entropy is achieved by the maximally mixed state, and equals $\log(\dim(U))$. Therefore $\text{scap}(A) = 2^{\log(\dim(U))} = \dim(U)$. Now let U be infinite dimensional. Then there is no upper bound on the entropy, and therefore $\text{scap}(A) = \infty = \dim(U)$. □

1.8 Fraction capacity

In this section we are going to define the **fraction capacity**, which can be understood as a generalization of probability that works with any ensemble space. In classical probability, additivity is justified by the mutual exclusivity of all elements of the sample space. In quantum mechanics, classical probability can be recovered only on orthogonal subspaces precisely because they represent mutually exclusive events. Therefore, the fraction capacity achieves the generalization not by focusing on probability of outcomes, but by focusing on mixing coefficient.

The typical way to understand probability is through outcomes of a process: 50% probability for tails means that if we repeated the coin toss multiple times, we would expect roughly half to be tails. In classical mechanics, it can also be understood as the probability of preparation: roughly half the times we selected a preparation procedure that prepared tails. In quantum mechanics, this does not work as the probability used during the mixing is the same as the probability of the outcome only if we are mixing orthogonal states. Effectively, probability in the usual sense is defined only on outcomes. The fraction capacity, instead, defines a measure on preparations.

The fraction capacity tells us how much of an ensemble e can be constructed through a mixture of ensembles from a set A. It is a non-negative, unit bounded, subadditive, monotone continuous (from below) measure, and it reduces to the probability measure in the classical case, and over measurement contexts in the quantum case. The goal is to create a measure theoretic generalization of probability theory that can work on all ensemble spaces.

First we define the **fraction** of one element with respect to another. For example, as shown in Fig. 1.6, suppose e_{123456} is a uniform distribution for a six-faced die. Suppose e_3 represents the outcome 3 with 100% probability. Then e_{123456} can be understood as $\mathsf{e}_{123456} = \frac{1}{6}\mathsf{e}_3 + \frac{5}{6}\mathsf{e}_{12456}$ where e_{12456} is the uniform distribution over the outcomes 1,2,4,5 and 6. Note that $\frac{1}{6}$ is the

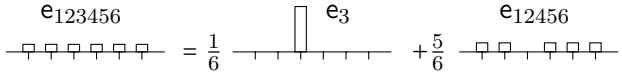
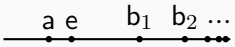


Figure 1.6: Visual representation of the fraction.

highest coefficient we can put in front of e_3 in a convex combination and have e_{123456} as a result, which coincides with the probability of obtaining 3 from a uniform distribution over six outcomes. That is how we define the fraction of e_3 with respect to e_{123456}.

Definition 1.170. *Let* $\mathsf{e}, \mathsf{a} \in \mathcal{E}$ *be two ensembles. The* ***fraction*** *of* a *with respect to* e *is the greatest mixing coefficient for which* e *can be expressed as a mixture of* a*. That is,* $\mathrm{frac}_{\mathsf{e}}(\mathsf{a}) = \sup(\{p \in [0,1] \,|\, \exists\, \mathsf{b} \in \mathcal{E} \text{ s.t. } \mathsf{e} = p\mathsf{a} + \bar{p}\mathsf{b}\})$.



Remark. We need to take the supremum as the maximum may not exist. For example, as shown in the figure, consider a discrete classical ensemble space and remove the extreme points. At the moment, there is no reason to rule out such space as unphysical.

Corollary 1.171. *Let* $\mathsf{e}, \mathsf{a} \in \mathcal{E}$*, then* $\mathrm{frac}_{\mathsf{e}}(\mathsf{a}) = 0$ *if and only if* a *is not a component of* e*.*

Proof. Note that a is a component of e if we can write $\mathsf{e} = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b}$ for some $\lambda \in (0,1]$ and $\mathsf{b} \in \mathcal{E}$. In this case, $\mathrm{frac}_{\mathsf{e}}(\mathsf{a}) \geq \lambda > 0$. Conversely, if $\mathrm{frac}_{\mathsf{e}}(\mathsf{a}) > 0$ we can find $\lambda \in (0, \mathrm{frac}_{\mathsf{e}}(\mathsf{a})]$ such that $\mathsf{e} = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b}$ for some $\mathsf{b} \in \mathcal{E}$. □

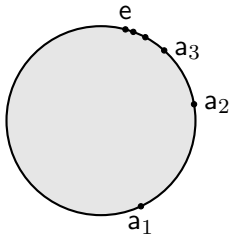


Figure 1.7: Discontinuity of fraction.

Intuitively, we would expect the fraction $\mathrm{frac}_{\mathsf{e}}(\mathsf{a})$ to be continuous both in a and e, but this is not the case. For example, as shown in Fig. 1.7, let $\mathcal{E}$ be the Bloch ball and let a_i be a sequence of pure states (i.e. extreme points) that converge to another pure state e. Since they are all extreme points, we have $\mathrm{frac}_{\mathsf{e}}(\mathsf{a}_i) = 0$ and $\mathrm{frac}_{\mathsf{a}_i}(\mathsf{e}) = 0$ for all i. However, $\mathrm{frac}_{\mathsf{e}}(\mathsf{e}) = 1$. This is the recurring problem of the non-continuity at the edges. The following conjecture, however, may still be true:

Conjecture 1.172. *The fraction* $\mathrm{frac}_{\mathsf{e}}(\mathsf{a})$ *is a continuous function both in* a *and* e *in the algebraic interior.*

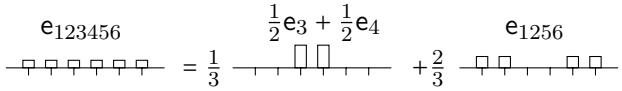


Figure 1.8: Visual representation of the fraction capacity.

We now define **fraction capacity** of a set of ensembles with respect to another ensemble. Like before, as shown in Fig. 1.8, suppose e_{123456} is a uniform distribution for a six-faced die, but now take $A = \{\mathsf{e}_3, \mathsf{e}_4\}$ as, respectively, the outcomes 3 and 4 with 100% probability. Then e can be understood as $\mathsf{e}_{123456} = \frac{1}{3}\left(\frac{1}{2}\mathsf{e}_3 + \frac{1}{2}\mathsf{e}_4\right) + \frac{2}{3}\mathsf{e}_{1256}$, where e_{1256} is the uniform distribution over outcomes 1,2,5 and 6. Again, note that $\frac{1}{3}$ is the highest coefficient we can put in front of any convex combination of elements of A, and still make a convex combination that has e as a result. That is how we define the fraction capacity of A with respect to e_{123456}.

The fraction capacity, then, defines how much of the ensemble e can be constructed with elements of A. The term capacity is used first because, intuitively, it tells us how much A can hold, and second because capacity is a name used to describe non-additive measures. The fraction capacity, in fact, has several nice properties. First, its value is always between zero and one, since the coefficient of a convex combination must be so bound. Second, it is monotone in the sense that if A gets bigger, the fraction capacity cannot decrease. Third, it is subadditive, meaning that the fraction capacity of the union of two sets must be the sum of the respective fraction capacities or less. Fourth, it is continuous from below. Note that if subadditivity is replaced by additivity, these are exactly the defining properties of a probability measure, since additivity and continuity are equivalent to σ-additivity.

Definition 1.173. *Let $\mathsf{e} \in \mathcal{E}$ be an ensemble and $A \subseteq \mathcal{E}$ a subset. The **fraction capacity** of A with respect to e is the biggest fraction achievable with convex combinations of A. That is,* $\mathrm{fcap}_{\mathsf{e}}(A) = \sup(\mathrm{frac}_{\mathsf{e}}(\mathrm{hull}(A)) \cup \{0\})$.

Corollary 1.174. *The fraction capacity is a convex supremum of the fraction with 0 for the empty set.*

Remark. Since we have shown that the fraction is not a continuous function, we can use the same example to show that the fraction capacity of the closed hull is not necessarily the same as the fraction capacity of the convex hull. As previously shown in Fig. 1.7, let $\mathcal{E}$ be a Bloch ball and let $A = \{\mathsf{a}_i\}$ be a sequence of pure states (i.e. extreme points) that converge to another pure state $\mathsf{e} \notin \{\mathsf{a}_i\}$. Since e is an extreme point, it can only be written as a convex combination of itself. Since $\mathsf{e} \notin A$, $\mathsf{e} \notin \mathrm{hull}(A)$ and therefore $\mathrm{fcap}_{\mathsf{e}}(\mathrm{hull}(A)) = 0$. However, since e is the limit of the sequence, we have $\mathsf{e} \in \mathrm{chull}(A)$, and therefore $\mathrm{fcap}_{\mathsf{e}}(\mathrm{chull}(A)) = 1$. The fraction capacity of the closed hull, then, can be different from the fraction capacity of the convex hull.

An open question is whether some type of equality still holds for internal points and the σ-hull of internal points. For example, $\mathrm{fcap}_{\mathsf{e}}(\mathrm{hull}(A)) = \mathrm{fcap}_{\mathsf{e}}(\mathrm{shull}(A))$ if e is an internal point and A is a set of internal points.

Corollary 1.175. *The fraction capacity uniquely identifies an ensemble. That is, let* $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ *such that* $\mathsf{a} \neq \mathsf{b}$. *Then* $\text{fcap}_{\mathsf{a}} \neq \text{fcap}_{\mathsf{b}}$.

Proof. Note that $\mathsf{e} = 1\mathsf{a} + 0\mathsf{b}$ if and only if $\mathsf{e} = \mathsf{a}$. Therefore, let $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ such that $\mathsf{a} \neq \mathsf{b}$. We have $\text{fcap}_{\mathsf{a}}(\{\mathsf{a}\}) = 1$ and $\text{fcap}_{\mathsf{b}}(\{\mathsf{a}\}) \neq 1$. Which means $\text{fcap}_{\mathsf{a}} \neq \text{fcap}_{\mathsf{b}}$. □

Proposition 1.176. *The fraction capacity with respect to an ensemble is a set function that is*

1. *non-negative and unit bounded:* $\text{fcap}_{\mathsf{e}}(A) \in [0, 1]$
2. *increasing:* $A \subseteq B \implies \text{fcap}_{\mathsf{e}}(A) \leq \text{fcap}_{\mathsf{e}}(B)$
3. *subadditive:* $\text{fcap}_{\mathsf{e}}(A \cup B) \leq \text{fcap}_{\mathsf{e}}(A) + \text{fcap}_{\mathsf{e}}(B)$
4. *continuous from below:* $\text{fcap}_{\mathsf{e}}(\lim_{i\to\infty} A_i) = \lim_{i\to\infty} \text{fcap}_{\mathsf{e}}(A_i)$ *for any increasing sequence* $\{A_i\}$

Proof. 1. By 1.51, $\text{fcap}_{\mathsf{e}}(A) \in [0, \sup(\text{frac}_{\mathsf{e}}(\mathcal{E}))] = [0, 1]$.

2. By 1.51.

3. Let $A, B \subseteq \mathcal{E}$ and let $p \in [0, 1]$ such that $\mathsf{e} = p\mathsf{e}_1 + \bar{p}\mathsf{e}_2$ for some $\mathsf{e}_1 \in \text{hull}(A \cup B)$ and $\mathsf{e}_2 \in \mathcal{E}$. Since $\mathsf{e}_1 \in \text{hull}(A \cup B)$, we can write $\mathsf{e}_1 = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b}$ for some $\lambda \in [0, 1]$, $\mathsf{a} \in A$ and $\mathsf{b} \in B$. Therefore we have $\mathsf{e} = p\lambda\mathsf{a} + p\bar{\lambda}\mathsf{b} + \bar{p}\mathsf{e}_2$. By the definition of fraction capacity, we must have $p\lambda \leq \text{fcap}_{\mathsf{e}}(A)$ and $p\bar{\lambda} \leq \text{fcap}_{\mathsf{e}}(B)$, therefore $p = p\lambda + p\bar{\lambda} \leq \text{fcap}_{\mathsf{e}}(A) + \text{fcap}_{\mathsf{e}}(B)$. Since $\text{fcap}_{\mathsf{e}}(A \cup B)$ is the supremum for a set of coefficients p for which the expression always holds, we have $\text{fcap}_{\mathsf{e}}(A \cup B) \leq \text{fcap}_{\mathsf{e}}(A) + \text{fcap}_{\mathsf{e}}(B)$. The fraction capacity is subadditive.

4. By 1.51. □

Remark. The fraction capacity is, in general, not continuous from above. Let $\mathcal{E}$ be the Bloch ball. Let $\mathsf{e} \in \mathcal{E}$ be the maximally mixed state (i.e. the center of the sphere). Let $\{\mathsf{e}_i\}$ be a countable set of distinct pure states. Let $A_j = \{\mathsf{e}_i \mid i \geq j\}$. Then $\text{fcap}_{\mathsf{e}}(A_j) \geq \frac{1}{2}$ for all j because A_j contains at least one pure state. This means that $\lim_{i\to\infty} \text{fcap}_{\mathsf{e}}(A_i) \geq \frac{1}{2}$. However, $\bigcap A_j = \varnothing$ which means $\text{fcap}_{\mathsf{e}}(\lim_{i\to\infty} A_i) = 0$. This means that, in general, $\text{fcap}_{\mathsf{e}}(\lim_{i\to\infty} A_i) \neq \lim_{i\to\infty} \text{fcap}_{\mathsf{e}}(A_i)$ for a decreasing sequence.

One open question is whether subadditivity extends to the infinite case. Since additivity plus continuity implies σ-additivity[14], it may be true that subadditivity plus continuity implies σ-subadditivity.

Conjecture 1.177. *The fraction capacity is* σ*-subadditive. That is,* $\text{fcap}_{\mathsf{e}}(\bigcup A_i) \leq \sum_i \text{fcap}_{\mathsf{e}}(A_i)$.

It is still an open question to understand exactly when the fraction capacity recovers additivity. We believe that this should be related to the ability to write an element as a mixture of separate components.

Conjecture 1.178. *The fraction capacity* fcap_{e} *is additive over* $A, B \subseteq \mathcal{E}$ *if and only if there exists a* $C \subseteq \mathcal{E}$ *such that* $A \Perp B$, $B \Perp C$, $A \Perp C$ *and* $\mathsf{e} \in \text{hull}(A \cup B \cup C)$.

We now show that fraction capacity recovers classical probability in the classical case. The key insight is that each event, each Borel set A, corresponds to a subspace of probability

[14] See for example Michel Grabisch, Set Functions, Games and Capacities in Decision Making.

measures for which that event will be true. This is the set of probability measure whose support is always within the given Borel set A. A probability measure, then, can always be decomposed in a part for which that event will be true, the part supported by A, and a part for which that event will be false, the part supported by the complement. The fraction capacity corresponds to the mixing coefficient of this decomposition.

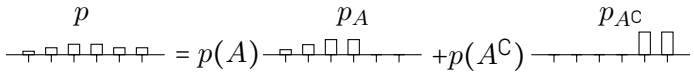


Figure 1.9: Connection between probability and fraction capacity.

Proposition 1.179. *Let $\mathcal{E}$ be a discrete or continuous classical ensemble space with sample space X. Let $A \in \Sigma_X$ be an event and $\mathcal{E}_A \subseteq \mathcal{E}$ the set of distributions with support contained in A. Let $p \in \mathcal{E}$ a probability measure in the ensemble space. Then $p(A) = \mathrm{fcap}_p(\mathcal{E}_A)$. That is, the probability of an event is equal to the fraction capacity of the subspace corresponding to that event.*

Proof. Let $p \in \mathcal{E}$ be a probability measure over X, absolutely continuous with respect to the appropriate measure μ (i.e. discrete for the discrete case and Liouville for the continuous case). Given $A \in \Sigma_X$, we can write $p = p(A)p_A + p(A^\complement)p_{A^\complement}$ where p_A is a probability measure with support with A and $p_{A^\complement}$ is a probability measure with support with $A^\complement$.

Let $\mathcal{E}_A \subseteq \mathcal{E}$ be the subset of probability measures with support contained in A. Then $p_A \in \mathcal{E}_A$ and $\mathrm{fcap}_p(\mathcal{E}_A) \geq p(A)$. Given that $p_{A^\complement}$ has support disjoint from A, we can find no further component of p that has support in A. Therefore $\mathrm{fcap}_p(\mathcal{E}_A) = p(A)$. The fraction capacity, then, recovers the probability. □

We now show that the fraction capacity recovers probability over measurements in the quantum case. The setup is essentially the same that as of the classical case, except that we start with an observable O, and each Borel set A for that observable will have a corresponding subspace of $\mathcal{H}$ and a projector Π_A. A post-measurement state of O is a mixed state that commutes with O, since it will commute with any projector that commutes with O. We can show, then, that the fraction capacity recovers the decomposition of any post-measurement state of O over the subspace corresponding to Π_A.

Proposition 1.180. *Let $\mathcal{E}$ be a quantum ensemble space associated to a Hilbert space $\mathcal{H}$. Let $O : \mathcal{H} \to \mathcal{H}$ be a quantum observable. Let $A \in \Sigma_{\mathbb{R}}$ be a Borel set and Π_A be the projector associated with the subspace of $\mathcal{H}$ corresponding to the subset A of the spectrum of O. Let $\mathcal{E}_A$ be the set of ensembles within that subspace. Let $\rho \in \mathcal{E}$ be a mixed state that commutes with O (i.e. the ensemble resulting from a measurement of O). Then $\mathrm{tr}(\rho\Pi_A) = \mathrm{fcap}_\rho(\mathcal{E}_A)$.*

Proof. Let $\rho \in \mathcal{E}$ be a mixed state that commutes with O. Let $A \in \Sigma_{\mathbb{R}}$ be a Borel set and Π_A be the projector associated with the subspace of $\mathcal{H}$ corresponding to the subset A of the spectrum of O. Then we can write $\rho = \lambda\rho_A + \bar{\lambda}\rho_{A^\complement}$ where $\rho_A, \rho_{A^\complement} \in \mathcal{E}$ are such that $\mathrm{tr}(\rho_A\Pi_A) = 1$ and $\mathrm{tr}(\rho_{A^\complement}\Pi_A) = 0$. We have $\mathrm{tr}(\rho\Pi_A) = \lambda$.

Let $\mathcal{E}_A \subseteq \mathcal{E}$ be the subset of mixed states d such that $\mathrm{tr}(d\Pi_A) = 1$. Then $d \in \mathcal{E}_A$

and $\text{fcap}_\rho(\mathcal{E}_A) \geq \lambda$. Given that $\rho_{A^\complement}$ is in the subspace orthogonal to $\mathcal{E}_A$, we can find no further component of ρ in that subspace. Therefore $\text{fcap}_\rho(\mathcal{E}_A) = \lambda$. We have $\text{tr}(\rho\Pi_A) = \lambda = \text{fcap}_\rho(\mathcal{E}_A)$ The fraction capacity, then, recovers the probability. □

1.9 Statistical properties and quantities

In this section we define statistical quantities, which are the generalization of classical random variables and quantum observables. Given our ensemble-first approach, statistical quantities are affine (linear) functions of ensembles as they represent an expectation over ensembles. The possible values of the quantities will need to be recovered with constructions that mimic spectral theory, though we will not want to talk about eigenstates and eigenvalues in the general case since these are not properly defined for a continuous quantity.

Properties and quantities

These definitions extend the notion of properties and quantities that we defined for an experimental domain. In that case, we simply required that a property be a continuous map to a set of possible values for the property. For a statistical property, the set of values are a topological convex space, so that the property allows statistical mixing. For a quantity, the topological convex space is simply a linearly ordered property, which in this case will always be a real-valued quantity. The convexity, the averaging operation, fills in the gaps in the linear order in case that all intermediate values can be obtained. For example, even if the values on the pure states are rationals or integers, the expectations will always span a contiguous interval of the reals.

Definition 1.181. *A **statistical property**, or simply property, is an attribute that allows statistical mixing. Formally, it is a continuous map $F : \mathcal{E} \to \mathcal{Q}$ where $\mathcal{Q}$ is a convex topological space such that $F(p\mathsf{a} + \bar{p}\mathsf{b}) = pF(\mathsf{a}) + \bar{p}F(\mathsf{b})$ (i.e. it is an affine map).*

*A **statistical quantity**, or statistical variable, or simply variable, is a numerical statistical property. That is, it is a continuous real-valued affine map $F : \mathcal{E} \to \mathbb{R}$.*

Justification. This definition extends to the statistical case the general definition of properties and quantities we already gave in the previous chapter. As before, continuity is required since verifying the value of the quantity corresponds to verifying that we are dealing with a specific subset of ensembles.

The ability to create convex combinations corresponds to the ability to create statistical averages. Therefore $\mathcal{Q}$ must be a convex set. Given that preparation instances are assumed to be independent, a mixture of the preparations will produce a mixture of the properties according to the same fraction coefficients.

Quantities are simply properties that are linearly ordered. The ability to create convex combinations becomes the ability to take weighted averages of the quantities. Regardless of whether one starts from integers, rationals or real-valued quantities, the statistical averages will, in general, be a real number.

Note that the numerical value cannot be infinite. Given that we can only measure the finite average of a finite sequence of outcomes, we must have that these sampled averages converge to the expectation. The only way this could happen for an infinite expectation would be if the sampled average kept increasing over time. But this would be a contradiction

of the assumption that an ensemble represents a reproducible collection of preparations. Therefore, statistical variables with infinite expectations are not justified. □

Remark. Note that variables will have a contiguous range on the ensembles because, given two ensembles with different values, we can mix them to obtain any intermediate value. This does not mean that the variable can take all possible values on pure states. For example, the number of particles in a pure state will necessarily be a non-negative integer, but we can mix those to create ensembles that have non-integer average number of particles.

Corollary 1.182. *A set of statistical quantities $\{F_i\}_{i\in I}$ can be collected into a single statistical property $F : \mathcal{E} \to \mathbb{R}^I$ where $F(\mathsf{e}) = \{F_i(\mathsf{e})\}_{i\in I}$.*

Here we show that statistical quantities recover the standard random variables of classical mechanics. That is, every statistical variable is a random variable and every random variable is a statistical variable.

Proposition 1.183. *Let $\mathcal{E}$ be a discrete or continuous classical ensemble space. Then each statistical quantity is the expectation of a random variable and vice-versa.*

Proof. Let $\mathcal{E}$ be a discrete or continuous classical ensemble space over some sample space X. Let $f : X \to \mathbb{R}$ be a random variable and $p \in \mathcal{E}$ a probability measure, then the expectation $F(p) \mapsto \int_X f dp$ is a continuous real affine map and is therefore a statistical quantity. Conversely, let $F : \mathcal{E} \to \mathbb{R}$ be a linear functional of the measures. Then, by the Riesz representation theorem, we can write $F(p) = \int_X f dp$ for some $f : X \to \mathbb{R}$. □

For a quantum system, we can show that all expectations of observables are statistical quantities.

Proposition 1.184. *Let $\mathcal{E}$ be a quantum ensemble space. All expectations of observables are statistical quantities.*

Proof. Let $\mathcal{E}$ be a quantum ensemble space over some Hilbert space $\mathcal{H}$. Let $O : \mathcal{H} \to \mathcal{H}$ be a Hermitian operator and ρ be a density matrix. Then $F(\rho) \mapsto \mathrm{tr}(\rho O)$ is a continuous real affine map of $\mathcal{E}$ and it is therefore a statistical quantity. □

Remark. We still need to prove that the converse is true: every statistical quantity in quantum mechanics corresponds to an operator. We may need to constrain the problem to bounded quantities, and recover the unbounded ones only as a limit.

Quantifiable spaces and locally convex vector spaces

In physics, we typically are able to fully identify ensembles by using measurable quantities. An ensemble space is quantifiable, then, when we have enough statistical quantities to recover all ensembles. There is a subtle problem with infinity, though, that is still not completely closed.

In principle, the space of ensembles may include quantities whose expectation is infinite over some distributions. In classical mechanics, for example, given any probability distribution ρ with infinite support we can find a random variable f such that $\int_X f\rho dx$ diverges. For example, we can simply set $f(x) = 1/\rho(x)$ where $\rho(x) \neq 0$ and $f(x) = 0$ otherwise. This can be solved in classical mechanics because we can always find distributions with compact

support, which will guarantee convergence for all functions. In quantum mechanics, however, a wavefunction with compact support in position will have non-compact support in momentum and vice-versa. However, if we were to restrict ourselves to the Schwartz wavefunctions we would have finite expectation for all polynomials of position and momentum, and they would be dense in the Hilbert space. What makes sense to expect physically, then, is that there is a set of quantities that physically define the system which will need finite expectation. These will need to remain finite under coordinate transformations and time evolution for the physics to make sense. All these details are yet to be understood and framed in a mathematically precise way. Yet, it is clear that we need, at least, to characterize the case where a set of statistical quantities fully characterizes all ensembles.

Definition 1.185. *A **quantifiable** ensemble space is an ensemble space where each ensemble can be identified by a set of statistical quantities. That is, there is family of statistical variables $F_i : \mathcal{E} \to \mathbb{R}$ such that, given $\mathsf{e}_1, \mathsf{e}_2 \in \mathcal{E}$, $F_i(\mathsf{e}_1) = F_i(\mathsf{e}_2)$ for all i if and only if $\mathsf{e}_1 = \mathsf{e}_2$. Moreover, the topology is generated by those quantities.*

We are going to show that the statistical quantities can be used to construct seminorms on the embedding vector space inducing a topology that is Hausdorff and locally convex. Since the statistical quantities are continuous in both the topology of the ensemble space and on this topology of the embedding vector space, the embedding of the ensemble space is continuous. Note that this embedding is not necessarily a homeomorphism onto the image as convergence of the expectation of continuous variables (i.e. weak convergence) does not guarantee convergence of the entropy (which requires, for example, convergence of the related inner products in the classical and quantum cases). Therefore, all the topological details of the ensemble space remain open even in this case.

Proposition 1.186. *A statistical variable F induces a seminorm on the vector space that embeds $\mathcal{E}$.*

Proof. Let F be a variable on $\mathcal{E}$, let $\mathsf{a} \in \mathcal{E}$ be an internal point and let V_a be the vector space of differences. Define $F_\mathsf{a} : V_\mathsf{a} \to \mathbb{R}$ such that $F_\mathsf{a}([r(\mathsf{b}-\mathsf{a})]) = |r(F(\mathsf{b}) - F(\mathsf{a}))|$.

First we show that F_a does not depend on the representative. If we have the zero class, then either $r = 0$ or $\mathsf{b} = \mathsf{a}$. In both cases, the function evaluates to zero. For the non-zero class, we have

$$\begin{aligned} F_\mathsf{a}\left(\left\{(r+j)\left(\left(\frac{r}{r+j}\mathsf{b} + \frac{j}{r+j}\mathsf{a}\right) - \mathsf{a}\right)\right\}\right) &= \left\{\left|(r+j)\left(F\left(\frac{r}{r+j}\mathsf{b} + \frac{j}{r+j}\mathsf{a}\right) - F(\mathsf{a})\right)\right|\right\} \\ &= \left\{\left|(r+j)\left(\frac{r}{r+j}F(\mathsf{b}) + \frac{j}{r+j}F(\mathsf{a}) - F(\mathsf{a})\right)\right|\right\} \\ &= \{|rF(\mathsf{b}) + jF(\mathsf{a}) - (r+j)F(\mathsf{a})|\} \\ &= \{|r(F(\mathsf{b}) - F(\mathsf{a}))|\}. \end{aligned} \tag{1.187}$$

Since F_a evaluates to an absolute value, it is a non-negative function. Since $F_\mathsf{a}(r[s(\mathsf{b}-\mathsf{a})]) = F_\mathsf{a}([(rs)(\mathsf{b}-\mathsf{a})]) = |(rs)(F(\mathsf{b}) - F(\mathsf{a}))| = |r||s(F(\mathsf{b}) - F(\mathsf{a}))| = |r|F_\mathsf{a}([s(\mathsf{b}-\mathsf{a})])$, F_a is

absolutely homogeneous. We also have

$$\begin{aligned}
&F_{\mathsf{a}}([r(\mathsf{b}-\mathsf{a})]+[s(\mathsf{c}-\mathsf{a})])\\
&= F_{\mathsf{a}}\left(\left[(r+s+k)\left(\left(\frac{r}{r+s+k}\mathsf{b}+\frac{s}{r+s+k}\mathsf{c}+\frac{k}{r+s+k}\mathsf{a}\right)-\mathsf{a}\right)\right]\right)\\
&= \left|(r+s+k)\left(F\left(\frac{r}{r+s+k}\mathsf{b}+\frac{s}{r+s+k}\mathsf{c}+\frac{k}{r+s+k}\mathsf{a}\right)-F(\mathsf{a})\right)\right|\\
&= |rF(\mathsf{b})+sF(\mathsf{c})+kF(\mathsf{a})-(r+s+k)F(\mathsf{a})|\\
&= |r(F(\mathsf{b})-F(\mathsf{a}))+s(F(\mathsf{c})-F(\mathsf{a}))|\\
&\leq |r(F(\mathsf{b})-F(\mathsf{a}))|+|s(F(\mathsf{c})-F(\mathsf{a}))|\\
&= F_{\mathsf{a}}([r(\mathsf{b}-\mathsf{a})])+F_{\mathsf{a}}([s(\mathsf{c}-\mathsf{a})]).
\end{aligned} \tag{1.188}$$

This means that F_{a} is subadditive. Therefore F_{a} is a seminorm by definition. □

Proposition 1.189. *A quantifiable ensemble space embeds continuously into a Hausdorff locally convex topological vector space.*

Proof. Let $\mathcal{E}$ be a quantifiable ensemble space and V_{a} be the embedding vector space corresponding to the internal point $\mathsf{a} \in \mathcal{E}$. Consider all the seminorms defined by the family of statistical variables that make $\mathcal{E}$ quantifiable. These will induce a topology on V_{a} making it a topological vector space (see Prop 2.2 in here). Since the topology is generated by seminorms, it is locally convex. Since the seminorms fully identify each point, the topology is Hausdorff (see Prop 2.6 in here).

Now consider $\mathcal{E}$ as embedded into V_{a}. If $F : \mathcal{E} \to \mathbb{R}$ is a statistical variable, then $|rF(\mathsf{b})-F(\mathsf{a})|$ is also a statistical variable since difference, scalar multiplication and absolute value are all continuous. This means that the open balls given by the seminorm on the vector space restricted to the ensemble space are open sets as they correspond to the open balls in the ensemble space. Moreover, since all elements of F_i, the statistical quantities that characterize the ensembles, are continuous in the ensemble space, a convergent sequence in the ensemble space will correspond to the convergence of all the seminorms in the vector space, thus convergence in the topology generated by said seminorms. The embedding, then, is continuous. □

At this point, we are not guaranteed that ensembles can be characterized by a countable family of statistical variables. The issue is that, while a second countable locally convex topology is generated by countably many seminorms, we know that the topology of the ensemble space is second countable but we do not know whether the subtopology generated by the statistical variables through the seminorms is second countable. Physically, it would mean that we have a way to experimentally distinguish between two ensembles (i.e. a second countable topology) but not through the expectation of statistical quantities. The entropy would, for example, provide other more stringent ways. It is not yet clear whether this is something that can indeed happen, and therefore a potential source of new physics, or we simply are missing a way to prove the result. Currently, the second option seems more likely.

If one is able to prove that the topology induced by the statistical variables is second countable, or if we simply impose that a countable family of statistical variables is enough to distinguish between ensembles, then the space is a metrizable second countable locally convex

topological vector space.

Proposition 1.190. *If a quantifiable ensemble space is fully determined by countably many statistical variables, then it embeds continuously into a metrizable second countable locally convex topological vector space.*

Proof. If a quantifiable ensemble space is characterized by countably many statistical variables, the topology induced on the vector space is second countable. A Hausdorff second countable topological vector space is metrizable. □

Remark. Note that we are missing completeness in terms of the seminorms to obtain a Fréchet space and it is unclear we are going to have it. For the classical case, suppose we restrict ourselves to all absolutely continuous probability measures with compact support. Expectations of all polynomials of position and momentum fully identify the measure. However, Dirac measures are also fully identified by those quantities, which can be achieved as the limit of a uniform distribution whose support shrinks to a point. That would give us a sequence of convergent expectations whose corresponding ensembles do not converge in the space.

Proposition 1.191. *Discrete/continuous classical ensemble spaces and quantum ensemble spaces are quantifiable.*

Proof. For discrete classical spaces, the expectation of the indicator of each extreme point defines a countable set of quantities that fully identifies the distribution.

For continuous classical spaces, note that $L^1(\mathbb{R}^{2n})$ is a Hausdorff second countable locally convex topological space.

For quantum ensemble spaces, the expectations of projectors define a family of statistical variables that can distinguish any pair of mixed states. □

An open problem is whether statistical quantities are required by ensemble spaces or they are an additional requirement. If we were able to show that ensemble spaces always embed continuously into a locally convex second countable topological vector space, one may use the seminorms to define quantities. Another approach is to investigate whether the existence of linear transformations parameterized by real quantities, such as time evolution, gives us statistical quantities through the generators of the transformations. It would also mean that, if time is not modeled as a real-valued parameter, then the ensemble space would have to change.

Connection to Choquet theory

We now want to establish a connection between ensemble spaces and Choquet theory. Specifically, we want to show that the fraction capacity restricted to the extreme points corresponds to the supremum of all possible measures that represent an ensemble. Moreover, the fraction capacity over the extreme points is additive exactly when the ensemble space is a Choquet simplex.

The goal of Choquet theory is to study how a point of a compact convex set can be represented by a probability measure over its extreme points. First, we need to define what it means to represent an ensemble with a probability measure. Let E be a convex subset of a locally convex topological vector space V and X the set of extreme points. We say that

$\mathsf{e} \in E$ is represented by a probability measure $p : \Sigma_E \to [0,1]$ if $F(\mathsf{e}) = \int_E F dp$ for every continuous linear functional F on V. Moreover, we say that p is supported by the extreme points if $p(X) = 1$. Given any probability measure p over the extreme points X, we can always find an $\mathsf{e} \in E$ that is represented by p. Therefore, we have a map from the set of probability measures to the set of ensembles. Choquet theory characterizes how the map works in the opposite direction.

The first question is whether each ensemble is represented by a probability measure. Choquet theory tells us that this is the case:

Theorem 1.192 (Choquet-Bishop-de Leeuw)**.** *Let E be a compact convex subset of a locally convex topological vector space V, and let $\mathsf{e} \in E$. Then there exists a probability measure p_{e} on E which represents e and is supported by the extreme points X.*

The second question is whether the representation is unique. Choquet theory has a clear answer.

Theorem 1.193 (Choquet uniqueness)**.** *Let E be a compact convex subset of a locally convex topological vector space V. Then E is a Choquet simplex if and only if for each $\mathsf{e} \in E$ there exists a unique measure p_{e} which represents e and is supported by the extreme points X.*

The definition of a Choquet simplex is rather involved, so it is omitted. Intuitively, it is the generalization of the finite-dimensional simplex. The key point is that E is a Choquet simplex if and only if the ensemble space does not allow multiple decompositions in terms of extreme points.

Note that there is a mismatch between ensemble spaces and the definition required by Choquet theory. While ensemble spaces are convex subsets of a vector space, they will not necessarily be compact. For example, in the classical continuous case, the space of absolutely continuous probability distributions is not compact, even if we restrict ourselves to a compact subset of phase space. The issue is that the extreme points, the Dirac distributions, are not ensembles. Moreover, ensemble spaces are equipped with an entropy, and therefore have more structure than required by Choquet theory.

However, the link we want to establish to Choquet theory is limited to the fraction capacity, which can be defined for any convex set. Therefore we are going to study the relationship between the fraction capacity in all cases where Choquet theory is applicable. If an ensemble space $\mathcal{E}$ is compact, the results will apply directly. If an ensemble space $\mathcal{E}$ admits some type of compactification E, like extending the space of ensembles to a compact set, then the results will apply to these extreme points that are not necessarily ensembles.

We first note that there is a connection between mixtures and the representation through a probability measure. We can, in fact, write $\mathsf{e} = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b}$ if and only if e can be represented by the probability measure $p(\{\mathsf{a}\}) = \lambda$, $p(\{\mathsf{b}\}) = \bar{\lambda}$, and $p(\mathcal{E} \setminus \{\mathsf{a}, \mathsf{b}\}) = 0$. Therefore, if M_{e} is the set of probability measures that represent e, the fraction $\mathrm{frac}_{\mathsf{e}}(\mathsf{a})$ is given by the maximum value over $\{\mathsf{a}\}$ that can be taken by one of those measures. That is, $\mathrm{frac}_{\mathsf{e}}(\mathsf{a}) = \sup\{p(\{\mathsf{a}\}) \,|\, p \in M_{\mathsf{e}}\}$.

Proposition 1.194. *Let E be a compact convex subset of a locally convex topological vector space V and define the fraction capacity as for an ensemble space. Let $\mathsf{e} \in E$ be a point and M_{e} be the set of measures that represent e. Then $\mathrm{frac}_{\mathsf{e}}(\mathsf{a}) = \sup\{p(\{\mathsf{a}\}) \,|\, p \in M_{\mathsf{e}}\}$ for all $\mathsf{a} \in E$ and $\mathrm{fcap}_{\mathsf{e}}(A) = \sup\{p(A) \,|\, p \in M_{\mathsf{e}}\}$ for all $A \subseteq E$.*

Proof. Let $\mathsf{e} \in E$ and suppose $\mathsf{e} = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b}$ with $\mathsf{a}, \mathsf{b} \in E$ and $\lambda \in [0,1]$. Then the measure p such that $p(\{\mathsf{a}\}) = \lambda$, $p(\{\mathsf{b}\}) = \bar{\lambda}$ and $p(E \setminus \{\mathsf{a}, \mathsf{b}\}) = 0$ represents e. In fact, for any statistical variable F we have $F(\mathsf{e}) = \lambda F(\mathsf{a}) + \bar{\lambda} F(\mathsf{b}) = \int_E F dp$. The converse is true as well. If $p \in M_{\mathsf{e}}$ such that $p(\{\mathsf{a}\}) = \lambda$, $p(\{\mathsf{b}\}) = \bar{\lambda}$ and $p(E \setminus \{\mathsf{a}, \mathsf{b}\}) = 0$, then $\mathsf{e} = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b}$. Since a Dirac measure at a point a represents a, the convex combination of two Dirac measures represents the convex combination of the two corresponding points. Now let $p \in M_{\mathsf{e}}$ such that $p(\{\mathsf{a}\}) \neq 1$. We can write it as $p = p(\{\mathsf{a}\})\delta_{\mathsf{a}} + p(E \setminus \{\mathsf{a}\}) \frac{p|_{E\setminus\{\mathsf{a}\}}}{p(E\setminus\{\mathsf{a}\})}$ where δ_{a} is the Dirac measure over a. But since every probability measure will represent an element of E, we can find $\mathsf{b} \in E$ that is represented by $\frac{p|_{E\setminus\{\mathsf{a}\}}}{p(E\setminus\{\mathsf{a}\})}$. Therefore there is $\hat{p} \in M_{\mathsf{e}}$ such that $\hat{p}(\{\mathsf{a}\}) = p(\{\mathsf{a}\})$, $\hat{p}(\{\mathsf{b}\}) = p(E \setminus \{\mathsf{a}\})$ and $\hat{p}(E \setminus \{\mathsf{a}, \mathsf{b}\}) = 0$. Which means $\mathrm{frac}_{\mathsf{e}}(\mathsf{a}) = \sup(\{\lambda \in [0,1] \,|\, \exists\, \mathsf{b} \in E \text{ s.t. } \mathsf{e} = \lambda\mathsf{a} + \bar{\lambda}\mathsf{b}\}) = \sup\{p(\{\mathsf{a}\}) \,|\, p \in M_{\mathsf{e}}\}$.

Now let $A \subseteq E$ be convex and let M_A be the set of all measures whose support is in A. Each measure in M_A will represent a point in A. Also, since the expectation of a convex combination of measures corresponds to a convex combination of the expectation values, M_A will be convex as well. If p represents e and $p(A) \neq 0$, then $\frac{p|_A}{p(A)}$ is a probability measure that represents an element $\mathsf{a} \in A$ such that $\mathsf{e} = p(A)\mathsf{a} + p(A^{\mathsf{C}})\mathsf{b}$. Therefore $\mathrm{frac}_{\mathsf{e}}(A) = \sup\{p(A) \,|\, p \in M_{\mathsf{e}}\}$. □

If we restrict the fraction capacity to the set X of extreme points, it will represent the supremum of all probability measures over the extreme points. If E is a Choquet simplex, then each ensemble will be represented by a unique measure over the extreme points, therefore the fraction capacity over the extreme points will coincide with that measure, and will be additive. Conversely, if the fraction capacity over the extreme points is additive, the fraction capacity must be the supremum over a single measure. If the fraction capacity over the extreme points is additive for all ensembles, then each ensemble is represented by a unique probability measure over the extreme points. In other words, the additivity of the fraction capacity over the extreme points exactly captures the single decomposition of ensembles into pure states and the Choquet simplex case.

Proposition 1.195. *Let E be a compact convex subset of a locally convex topological vector space V. Then the following statements are equivalent:*

1. *E is a Choquet simplex*
2. *each ensemble is uniquely represented by a measure over its extreme points X*
3. *the fraction capacity restricted over the extreme points is additive*

Proof. By theorem 1.193, 1 and 2 are equivalent.

Now consider $\mathrm{frac}_{\mathsf{e}}|_X$. By 1.194, $\mathrm{fcap}_{\mathsf{e}}|_X$ gives us the supremum of all possible measures supported by X that represent e. By 1.192, e is represented by at least one measure. Suppose that it is represented by exactly one measure p. Then $\mathrm{frac}_{\mathsf{e}}|_X = p$ and therefore $\mathrm{frac}_{\mathsf{e}}|_X$ is

additive. Now, suppose that e is represented by at least two distinct measures p_1 and p_2. Then there will be a set $A \subseteq X$ such that $p_1(A) \neq p_2(A)$. Suppose, without loss of generality, that $p_1(A) > p_2(A)$. We will also have $p_1(A^\complement) = 1 - p_1(A) < 1 - p_2(A) = p_2(A^\complement)$. Therefore $\mathrm{fcap}_\mathsf{e}(A \cup A^\complement) = \mathrm{fcap}_\mathsf{e}(X) = 1 = p_1(A) + p_1(A^\complement) < p_1(A) + p_2(A^\complement) \leq \mathrm{fcap}_\mathsf{e}(A) + \mathrm{fcap}_\mathsf{e}(A^\complement)$. The fraction capacity is therefore additive if and only if e is represented by a single measure over the extreme points. Therefore 2 and 3 are equivalent. □

Macrostates and thermodynamics

In this section we try to recover some elements of thermodynamics on the generalized ensemble space. We want to recover Gibbs' thermodynamics, which means an equation of state in terms of extensive quantities. Instead of extensive quantities, we are going to use statistical quantities. Note that the idea that all extensive quantities are statistical (i.e. the average during mixing) seems to work. In fact, in statistical mechanics the energy, the number of particles and the volume are averages over microstates. It also seems that intensive quantities do not average during mixing. Temperature, for example, is only defined on equilibria (i.e. of Boltzmann distributions) and the mixture of two equilibria at different temperatures is not an equilibrium. Still, we would need a general proof, which would require the notion of product spaces (i.e. intensive/extensive quantities represent system/subsystem relationship).

Definition 1.196. *Let $\mathcal{E}$ be an ensemble space. Let $F : \mathcal{E} \to \mathcal{Q}$ be a statistical property. The **coarse graining** of $\mathcal{E}$ over F is the set $\mathcal{M} \subseteq \mathcal{E}$ represented by the ensembles that maximize the entropy for each fixed value of the property. That is, there exists a map $\psi : \mathcal{Q} \to \mathcal{M}$ such that $F(\psi(x)) = x$ and $S(\psi(x)) \geq S(\mathsf{e})$ for all $\mathsf{e} \in \mathcal{E}$ such that $F(\mathsf{e}) = x$. The **equation of state** is the map $S : \mathcal{Q} \to \mathbb{R}$ defined as $S(x) \mapsto S(\psi(x))$ that returns the entropy given the value of the statistical property.*

The set of ensembles that maximize entropy will be a set of points fully identified by finitely many real numbers. We would expect this to be a manifold, though we still have to understand how the topology of the ensemble space relates to the topology of $\mathcal{M}$.

Conjecture 1.197. *Let $F : \mathcal{E} \to \mathbb{R}^n$ be a vector of statistical quantities. Then the corresponding coarse graining $\mathcal{M}$ of $\mathcal{E}$ is a manifold. The equation of state $S : \mathbb{R}^n \to \mathbb{R}$ returns the entropy as a function of the values of the statistical quantities.*

Remark. We need to prove that $\mathcal{M}$ inherits the topology from $\mathcal{E}$. Are we be able to recover the topological isolation of phase transitions? The quantities will typically be energy plus other extensive quantities (i.e. volume, number of particles, ...).

What we can already prove, however, is that the entropy is a strictly concave function of the state variables. Note that, in this setting, the convex combination between state variables does not give a statistical mixture, but rather it gives us the ensemble that maximizes entropy on the averaged constraint. Therefore, the concavity of the entropy on the coarse grained state space is not the same concavity of the entropy on the ensemble space.

Proposition 1.198. *Given a coarse graining $\mathcal{M}$ of $\mathcal{E}$, its equation of state is strictly*

concave. That is, $S(\lambda x + \bar{\lambda} y) \geq \lambda S(x) + \bar{\lambda} S(y)$ *and the equality holds if and only if* $x = y$.

Proof. Let $x, y \in \mathcal{Q}$ be two possible values for the statistical property, and let $\lambda x + \bar{\lambda} y$ be a convex combination. We have $F(\psi(\lambda x + \bar{\lambda} y)) = \lambda x + \bar{\lambda} y = F(\lambda\psi(x) + \bar{\lambda}\psi(y))$. That is, $\psi(\lambda x + \bar{\lambda} y)$ and $\lambda\psi(x) + \bar{\lambda}\psi(y)$ are two ensembles that share the same value for the statistical property. By definition, the entropy of the first cannot be lower than the entropy of the second. We have

$$\begin{aligned} S(\lambda x + \bar{\lambda} y) &= S(\psi(\lambda x + \bar{\lambda} y)) \\ &\geq S(\lambda\psi(x) + \bar{\lambda}\psi(y)) \\ &\geq \lambda S(\psi(x)) + \bar{\lambda} S(\psi(y)) \\ &= \lambda S(x) + \bar{\lambda} S(y), \end{aligned} \tag{1.199}$$

which shows that the equation of state is concave.

To show that the equation of state is strictly concave, suppose that $x = y$. Then on one side $S(\lambda x + \bar{\lambda} y) = S(x)$ and on the other side $\lambda S(x) + \bar{\lambda} S(y) = S(x)$. Therefore $S(\lambda x + \bar{\lambda} y) = \lambda S(x) + \bar{\lambda} S(y)$, which means the inequality holds with the equal. Conversely, suppose that $x \neq y$. Then $\psi(x) \neq \psi(y)$. By the strict concavity of the entropy, we have

$$\begin{aligned} S(\lambda x + \bar{\lambda} y) &= S(\psi(\lambda x + \bar{\lambda} y)) \\ &\geq S(\lambda\psi(x) + \bar{\lambda}\psi(y)) \\ &> \lambda S(\psi(x)) + \bar{\lambda} S(\psi(y)) \\ &= \lambda S(x) + \bar{\lambda} S(y), \end{aligned} \tag{1.200}$$

which shows that the equality holds if and only if $x = y$ and therefore the equation of state is strictly concave. □

Remark. In the case of thermodynamics, where the statistical property is a vector of statistical values, the equation of state will be concave in all arguments and in all combinations of arguments.

1.10 Orthogonal and separate subspaces

In this section we recover the notion of subspaces from orthogonality and separateness. Since orthogonality as defined from the entropy coincides with the orthogonality of the inner product for classical and quantum ensemble spaces, orthogonal subspaces as defined here will coincide with those defined by the inner product.

Irreflexive symmetric relations and topped ∩-structures

We first present a generic construction, which is sometimes called orthogonality space or orthomodular space, and it sometimes appears in the context of Galois connections. The general idea is that given any irreflexive and symmetric relation, we can define a closure that gives us an orthocomplemented lattice. Since orthogonality and separateness, as defined before, satisfy those properties, they will each give a notion of subspace.

To get an idea, we look at how orthogonality and orthogonal subspaces work in inner product spaces. Let V be an inner product space. If we take a set $U \subseteq V$ of vectors, we can define the set $U^\perp$ of all the vectors that are orthogonal to all elements of U. This will return the subspace orthogonal to all elements of U. We can also define $(U^\perp)^\perp$ as the set of all the

vectors that are orthogonal to all elements of $U^\perp$. The set $(U^\perp)^\perp$ will contain all the elements of U, because they are all orthogonal to all elements of $U^\perp$ by definition, but it will also include all the elements in the same subspace. If U were a subspace to begin with, then $U = (U^\perp)^\perp$. Note that we can therefore define subspaces without having a notion of space, just by using the orthogonality relationship, by looking for sets such that $U = (U^\perp)^\perp$.[15]

The construction works for any binary relationship that is symmetric and irreflexive. For example, given two distributions, the fact that they have disjoint support is a symmetric and irreflexive relationship: a distribution does not have disjoint support with respect to itself, and the order of comparison does not matter. We can then construct subspaces based on this relationship, and find sets of functions that are all defined within different regions. In our case, separateness, as defined by the mixing function, and orthogonality, as defined by the entropy, are symmetric and irreflexive.

We start with a generic set X and a symmetric relation (i.e. if aRb then bRa). We define the R-complement U^R of all elements that are R-related to U and study some useful properties. The symmetry of the relation is already enough to recover most of the properties we will need.

Definition 1.201. *Let X be a set and $R \subseteq X \times X$ a symmetric relation. Given a subset $U \subseteq X$, we define the R-**complement** to be*

$$U^R = \{a \in X \mid \forall b \in U, aRb\}.$$

Proposition 1.202. *Let X be a set and $R \subseteq X \times X$ a symmetric relation. Then*

1. $U \subseteq V \implies V^R \subseteq U^R$
2. $U \subseteq (U^R)^R$
3. $U^R = ((U^R)^R)^R$
4. $U^R = (V^R)^R \iff (U^R)^R = V^R$
5. $(\bigcup_{i\in I} U_i)^R = \bigcap_{i\in I}(U_i)^R$
6. $\varnothing^R = X$

Proof. 1. Suppose $a \in V^R$. Then, by definition, $\forall b \in V, aRb$. Since $U \subseteq V$, it is also true that $\forall b \in U, aRb$. Therefore $a \in U^R$ by definition. Since a was arbitrary, $V^R \subseteq U^R$.

2. By expanding the definition of complement, we have $(U^R)^R = \{a \in X \mid \forall b \in U^R, aRb\} = \{a \in X \mid \forall b \in \{c \in X \mid \forall d \in U, cRd\}, aRb\} = \{a \in X \mid \forall b \in X \; s.t. (\forall d \in U, bRd), aRb\}$.

Let $a \in U$ and let $b \in X$ such that $\forall d \in U, bRd$. Since $a \in U$ and bRd for all $b \in U$, we have bRa in particular. Since R is symmetric, aRb. Given that b was arbitrary, we conclude that $\forall b \in X \; s.t. (\forall d \in U, bRd), aRb$. Therefore $a \in (U^R)^R$ by definition of complement. Given that a was arbitrary, $U \in (U^R)^R$.

3. We again expand the definition and have $((U^R)^R)^R = \{a \in X \mid \forall b \in (U^R)^R, aRb\}$.

Let $x \in ((U^R)^R)^R$. Then $\forall b \in (U^R)^R, xRb$ by definition of the complement. Since by 1. $U \subset (U^R)^R$, we can restrict the previous expression to only the elements of U, and therefore $\forall b \in U, xRb$. But this means that $x \in U^R$ by definition of the complement. Since x was arbitrary, $((U^R)^R)^R \subseteq U^R$. But by 1., we also have $U^R \subseteq ((U^R)^R)^R$ since U^R is just

[15] This type of construction is similar to some constructions related to Galois connections.

a set onto which we can apply the complement twice . By two-way containment, we have $U^R = ((U^R)^R)^R$.

4. Let $U, V \subseteq X$ such that $U^R = (V^R)^R$. Applying the complement on each side, $(U^R)^R = ((V^R)^R)^R$. By the previous property $((V^R)^R)^R = V^R$ and therefore $(U^R)^R = V^R$. Switching U and V proves the other direction.

5. We have:

$$\begin{aligned}
(\bigcup_{i\in I} U_i)^R &= \{a \in X \mid \forall b \in \bigcup_{i\in I} U_i, aRb\} \\
&= \{a \in X \mid \forall U_i, \forall b \in U_i, aRb\} \\
&= \{a \in X \mid \forall U_i, a \in \{c \in X \mid \forall b \in U_i, cRb\}\} \\
&= \bigcap_{i\in I} \{c \in X \mid \forall b \in U_i, cRb\} \\
&= \bigcap_{i\in I} (U_i)^R
\end{aligned}$$

Note that this property does not rely on the symmetry of R.

6. Let $a \in X$. There is no $b \in \varnothing$ such that aRb. Therefore $\forall b \in \varnothing, aRb$. This means that $a \in \varnothing^R$. Since a was arbitrary, $X = \varnothing^R$. □

The next two properties depend on both the symmetry and the irreflexivity.

Proposition 1.203. *Let X be a set and $R \subseteq X \times X$ a symmetric and irreflexive relation. Then*

1. $U \cap U^R = \varnothing$
2. $X^R = \varnothing$
3. $(U \cup U^R)^R = \varnothing$

Proof. 1. Let $a \in U$. Since R is irreflexive, aRa is false. Therefore it is not true that, for all $b \in U$, aRb. This means that $a \notin U^R$. Since a was arbitrary, $U \cap U^R = \varnothing$.

2. Suppose $a \in X^R$. Then for all $b \in X$, aRb. In particular, we would have aRa, which can't be true since R is irreflexive. Therefore $a \notin X^R$ and, since a is arbitrary, $X^R = \varnothing$.

3. Suppose $a \in (U \cup U^R)^R$. Then aRb for all $b \in U \cup U^R$. Since aRb for all $b \in U$, $a \in U^R$. But this would mean aRa, which is not possible since R is irreflexive. Therefore $(U \cup U^R)^R = \varnothing$. □

We now define a notion of R-subspace by requiring that a subspace is the R-complement of its R-complement. We then construct the lattice of subspaces and show it satisfies properties we would expect from a lattice of subspaces.

Definition 1.204. *Let X be a set and $R \subseteq X \times X$ an irreflexive symmetric relation. Let $U \subseteq X$. The **R-closure of** U is $\langle U \rangle_R = (U^R)^R$. An **$R$-subspace** of X is a set $U \subseteq X$ such that $U = \langle U \rangle_R$. The **lattice of R-subspaces** is the set $\mathfrak{L} = \{U \subseteq X \mid U = \langle U \rangle_R\}$ ordered by inclusion.*

Corollary 1.205. *The lattice of R-subspaces $\mathfrak{L}$ is a topped $\bigcap$-structure on X and therefore*

is also a complete lattice.

Proof. The set $\mathfrak{L}$ is a collection of subsets of X. Let $\{U_i\}_{i\in I} \subseteq \mathfrak{L}$ be a non-empty family. Then, using the definition of subspace, and the third and fifth properties of 1.202, we have

$$\begin{aligned}\bigcap_{i\in I} U_i &= \bigcap_{i\in I}(U_i^R)^R = (\bigcup_{i\in I} U_i^R)^R \\ &= ((((\bigcup_{i\in I}(U_i^R)^R)^R)^R = (((\bigcap_{i\in I}(U_i^R)^R)^R)^R \\ &= ((\bigcap_{i\in I} U_i)^R)^R.\end{aligned}$$

Therefore $\bigcap_{i\in I} U_i \in \mathfrak{L}$. This means $\mathfrak{L}$ is an $\bigcap$-structure. Using the second property of 1.203, $X = \varnothing^R = (X^R)^R$. Therefore $\mathfrak{L}$ is a topped $\bigcap$-structure. This also means that it is a complete lattice. □

Proposition 1.206. *The lattice of R-subspaces is orthocomplemented. That is*

1. *the R-complement is a lattice complement: $A \wedge A^R = \varnothing$ and $A \vee A^R = X$*
2. *$(A^R)^R = A$*
3. *$A \subseteq B$ implies $B^R \subseteq A^R$.*

Proof. Note that 2 is satisfied because A is an R-closure and 3 is already proven in 1.202. For 1, since the lattice is an $\bigcap$-structure, $A \wedge A^R = A \cap A^R$ and from 1.203 we have $A \cap A^R = \varnothing$. Since the lattice is an $\bigcap$-structure, $A \vee A^R \supseteq A \cup A^R$. Using 1.202 and 1.203 we have $((A \vee A^R)^R)^R \supseteq ((A \cup A^R)^R)^R = (\varnothing)^R = X$. □

Corollary 1.207. *The lattice of R-subspaces satisfies de Morgan's laws. That is*

1. *$(A \vee B)^R = A^R \wedge B^R$*
2. *$(A \wedge B)^R = A^R \vee B^R$*

Proof. Every orthocomplemented lattice satisfies de Morgan's laws.[a] □

Corollary 1.208. *The R-closure satisfies the following properties*

1. *$U \subseteq \langle U\rangle_R$*
2. *$U \subseteq V \implies \langle U\rangle_R \subseteq \langle V\rangle_R$*
3. *$\langle\langle U\rangle_R\rangle_R = \langle U\rangle_R$*

and is therefore a closure operation.

Proof. 1. The first property is true by the second property of 1.202.

2. Using the first property of 1.202 we have $U \subseteq V$ implies $V^R \subseteq U^R$ which in turn implies $(U^R)^R \subseteq (V^R)^R$. Therefore $\langle U\rangle_R \subseteq \langle V\rangle_R$.

3. Using the fifth property of 1.202 we have $\langle\langle U\rangle_R\rangle_R = (((U^R)^R)^R)^R = (U^R)^R = \langle U\rangle_R$. □

Proposition 1.209. *Let X be a set and $R \subseteq X \times X$ a symmetric and irreflexive relation. Then*

1. $\langle U\rangle_R$ *is the smallest* R*-subspace containing* U
2. *if* $U, V \in \mathfrak{L}$ *then* $U = V^R \iff U^R = V$

Proof. 1. Let $U \subseteq X$ and $V \in \mathfrak{L}$ such that $U \subseteq V$ and $V \subseteq \langle U\rangle_R$. Since $U \subseteq V$, using the second property of 1.208, $\langle U\rangle_R \subseteq \langle V\rangle_R = V$. Since $V \subseteq \langle U\rangle_R$ and $\langle U\rangle_R \subseteq V$, $\langle U\rangle_R = V$. This means that no R-subspace that contains U is smaller than $\langle U\rangle_R$.

2. Since $U, V \in \mathfrak{L}$, $U = (U^R)^R = V^R$ which, by the fourth property of 1.202, implies $U^R = (V^R)^R = V$. Switching U and V proves the other direction. □

[a] See, for example, Wikipedia.

We now show that if we start with the mere notion of orthogonality defined from an inner product vector space, we recover the notion of orthogonal components and subspaces. That is, the full structure of subspaces of an inner product space can be fully recovered only from pairwise orthogonality.

Proposition 1.210. *Let* X *be an inner product space and* $R = \{(a, b) \,|\, \langle a, b\rangle = 0\}$. *Then*

1. R *is an irreflexive and symmetric relation*
2. $U^R = U^\perp$
3. $\langle U\rangle_R = \mathrm{cl}_X(\mathrm{span}(U))$

Proof. 1. Given that the inner product is symmetric, so will be R. Given that no vector is orthogonal to itself, R is irreflexive.

2. The orthogonal complement is defined as $U^\perp = \{a \in X \,|\, \forall b \in U, \langle a, b\rangle = 0\}$. Since $aRb \iff \langle a, b\rangle = 0$, $U^R = U^\perp$.

3. We have $\langle U\rangle_R = (U^\perp)^\perp$ which returns the smallest closed subspace that contains U. □

Separate and orthogonal subspaces

Since orthogonality and separateness are irreflexive symmetric operations, they will both give a lattice of subspaces. Here we present a series of conjectures about these subspaces that we have yet to prove.

Definition 1.211. *Let* $\mathcal{E}$ *be an ensemble space. An* $\perp$*-subspace is an* R*-subspace defined by* $\perp$ *and a* $\between$*-subspace is an* R*-subspace defined by* $\between$.

In the classical case, since both orthogonality and separateness coincide with disjoint support, a subspace consists of all the probability distribution whose support is within a given set U.[16]

Proposition 1.212. *Let* $\mathcal{E}$ *be a discrete or continuous classical ensemble space over a sample space* X. *Then* $A \subseteq \mathcal{E}$ *is an* $\perp$*-subspace if and only if there exists* $U \subseteq X$ *such that* $A = \{p \in \mathcal{E} \,|\, p(U) = 1\}$.

[16] Clearly, not all closed sets will give a different subspace. For example, over $\mathbb{R}$, $[0, 1]$ and $[0, 1] \cup [2, 2]$ will correspond to the same subspace. Characterizing that equivalence class is still an open question.

Proof. Recall that in a classical ensemble space, $p \perp \lambda$ if and only if $\lambda(\mathrm{supp}(p)) = 0$. Let $A \subseteq \mathcal{E}$ be an $\perp$-subspace. Let $U = \bigcup_{p \in A} \mathrm{supp}(p)$ be the union of all supports. Then $\lambda \in A^\perp$ (i.e. is orthogonal to all elements of A) if and only if $\lambda(U) = 0$. Conversely, $p \in A$ only if it is orthogonal to all elements, which means if and only if $p(U) = 1$. Therefore if A is a subspace, there is a $U \subseteq X$ such that $A = \{p \in \mathcal{E} \,|\, p(U) = 1\}$.

Conversely, let $A = \{p \in \mathcal{E} \,|\, p(U) = 1\}$ for some $U \subseteq X$. Then $A^\perp = \{p \in \mathcal{E} \,|\, p(U) = 0\}$. The orthogonal complement of the orthogonal complement will return A, which means $A = (A^\perp)^\perp$ is a subspace. □

In the quantum case, the orthogonality coincides with orthogonality in the Hilbert space, and therefore an $\perp$-subspace coincides with the set of density operators supported by a given subspace.

Proposition 1.213. *Let $\mathcal{E}$ be a quantum ensemble space over a Hilbert space $\mathcal{H}$. Then $A \subseteq \mathcal{E}$ is an $\perp$-subspace if and only if there exists a subspace $U \subseteq \mathcal{H}$ with a corresponding projector $\mathbf{1}_U$ such that $A = \{\rho \in \mathcal{E} \,|\, \mathrm{tr}[\mathbf{1}_U \rho] = 1\}$.*

Proof. Recall that to each subspace $U \subseteq \mathcal{H}$ is associated a projector $\mathbf{1}_U$ whose eigenvectors with eigenvalue 1 span U. Given a density operator $\rho \in \mathcal{E}$, its support is the subspace $U = \mathrm{supp}(\rho)$ spanned by the non-zero eigenvectors. Therefore $\mathrm{tr}[\mathbf{1}_U \rho] = 1$ and $\mathrm{tr}[\mathbf{1}_{U^\perp} \rho] = 0$.

Also recall that, for a quantum ensemble space, $\rho \perp \sigma$ if and only if $\mathrm{tr}[\mathbf{1}_{\mathrm{supp}(\rho)} \sigma] = 0$. Let $A \subseteq \mathcal{E}$ be an $\perp$-subspace. Let $U = \bigvee_{\rho \in A} \mathrm{supp}(\rho)$ be the subspace U spanned by all supports. Then $\sigma \in A^\perp$ (i.e. is orthogonal to all elements of A) if and only if $\mathrm{tr}[\mathbf{1}_U \sigma] = 0$. Conversely, $\rho \in A$ if and only if it is orthogonal to all elements of $A^\perp$, which is the case if and only if $\mathrm{tr}[\mathbf{1}_U \rho] = 1$. Therefore if A is a subspace, there is a subspace $U \subseteq \mathcal{H}$ such that $A = \{\rho \in \mathcal{E} \,|\, \mathrm{tr}[\mathbf{1}_U \rho] = 1\}$.

Now let $A = \{\rho \in \mathcal{E} \,|\, \mathrm{tr}[\mathbf{1}_U \rho] = 1\}$ for some subspace $U \subseteq \mathcal{H}$. The orthogonal complement $A^\perp$ will be the set of all density operators $\sigma \in \mathcal{E}$ such that $\mathrm{tr}[\mathbf{1}_U \sigma] = 0$. The complement of the complement will return A, which means $A = (A^\perp)^\perp$ is a subspace. □

Note that, in general, the two notions of subspaces, and therefore their lattices, will be different. Note that a feature of classical ensemble spaces is exactly that orthogonality and separateness coincide, which also means the two lattices will coincide.

Another question is whether the lattice of $\perp$-subspaces has to be necessarily orthomodular, as one requires, for example, in quantum logic. The cut triangle 1.215 is a counterexample. The only orthogonality relationships are $2 \perp p3 + \bar{p}\mathsf{b}$ and $3 \perp p2 + \bar{p}\mathsf{a}$. This means that the only $\perp$-subspaces are the bottom $\perp = \varnothing$, $2 = \{2\}$, $3 = \{3\}$, $A2 = \{p2 + \bar{p}\mathsf{a}\}$, $B3 = \{p3 + \bar{p}\mathsf{b}\}$ and the top $\top = \mathcal{E}$. As we can is see in Fig. 1.10, the subspaces form an M_6 lattice, which is not orthomodular. In fact, $3 \subseteq B3$ but $3 \vee (3^\perp \wedge B3) = 3 \vee (A2 \wedge B3) = 3 \vee \perp = 3 \neq B3$. The lack of orthomodularity is due to the multiple complements where one is a subset of the other.

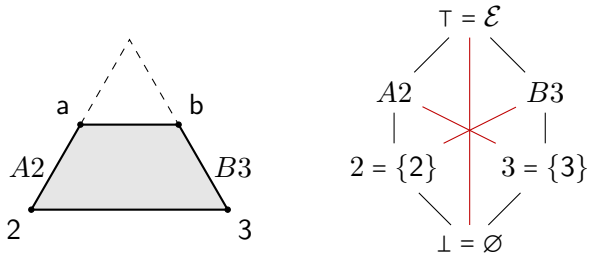


Figure 1.10: On the left, the cut triangle. On the right, the lattice of the orthogonal subspaces. The red lines connect the orthogonal complements.

This leads to the following:

Conjecture 1.214. *The lattice of $\perp$-subspaces is orthomodular if and only if the ensemble space is orthogonally decomposable (i.e. every decomposable ensemble is orthogonally decomposable).*

If this is true, the requirement of orthomodularity is an additional requirement that can be understood physically.

Note, however, that the lattice of π-subspaces of the cut triangle is orthomodular, though not commutative. This shows how the two lattices can be very different.

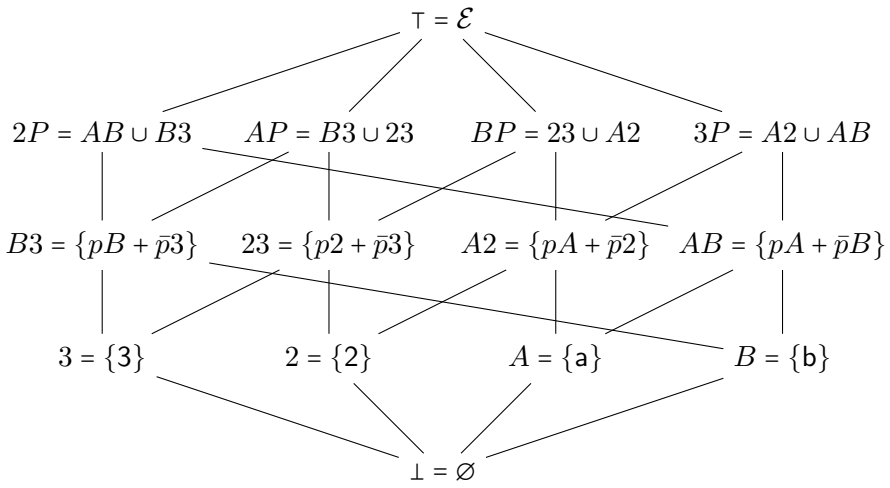


Figure 1.11: In this diagram, the lower tier represents the singletons of the extreme points, the middle tier represents each side and the top tier represents the pairs of sides that are the π-complement of the extreme points. The red lines connect the complements. Note that the lattice is not distributive (e.g. $A \vee (3 \wedge A2) = A \vee \perp = A \neq (A \vee 3) \wedge (A \vee A2) = BP \wedge A2 = A2$). In fact, $\{\perp, 3, A, A2, BP\}$ is an N_5 sublattice.

1.11 Examples

In this section we plan to collect examples that appear throughout the chapter to exemplify corner cases, wanted or unwanted.

Cut triangle

The cut triangle is an example of a simple classical space with a constraint, which makes it an interesting candidate for testing concepts and definitions on a small deviation from the classical case.

Example 1.215 (Cut triangle). The cut triangle is the subset of probability measures over three elements where the probability of the first is constrained to be no more than $\frac{1}{2}$.

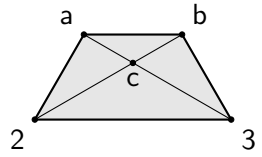


Figure 1.12: Cut triangle.

This ensemble space is the standard two simplex, a triangle, where the top is cut off. This is the space of probability distributions $[p_1, p_2, p_3]$ such that $p_1 \leq \frac{1}{2}$, where entropy is the Shannon entropy. The extreme points of the space are $\mathsf{2} = [0, 1, 0]$, $\mathsf{3} = [0, 0, 1]$, $\mathsf{a} = [\frac{1}{2}, \frac{1}{2}, 0]$ and $\mathsf{b} = [\frac{1}{2}, 0, \frac{1}{2}]$. Every ensemble that is decomposable is also separately decomposable but not necessarily orthogonally decomposable. Most ensembles, all those in the interior, are also separately multidecomposable. For example, c can be expressed as a mixture of a and $\mathsf{3}$, and of b and $\mathsf{2}$.

Real interval with right topology

This example shows how one can have a topological convex space that embeds in a vector space, but its topology is not compatible with the vector space operations. In a topological vector space, the multiplication by -1 forces sets that are the mirror image of open sets to be open. This is not guaranteed in a topological convex space.

Note that the topology in the example is T_0 but not T_1. The existence of an entropy function forces the space to be at least Hausdorff, therefore this is not a valid ensemble space.

Example 1.216 (Real interval with right topology). Let $\mathcal{E}$ be the interval $[0, 1] \subset \mathbb{R}$ with the topology generated by sets of the form $(r, 1]$. Since the topology is not T_1, it cannot be an ensemble space, but the mixing operation is continuous.

To show that the mixing operation is continuous, it suffices to show that the inverse image of sets of the form $(r, 1]$ is open. To do that, we show that each point in $+^{-1}((r, 1]) = \{(p, a, b) \in [0, 1]^3 \,|\, pa + \bar{p}b > r\}$ has an open neighborhood. First, note that sets of the form $(p, 1] \times (a, 1] \times (b, 1]$ and $[0, p) \times (a, 1] \times (b, 1]$ are open because they are the product of open sets. Next note that if $pa + \bar{p}b > m$, then if $a \geq b$, for any $\lambda > p$, $c > a$ and $d > b$, $\lambda c + \bar{\lambda} d > m$. This is because the convex combination can only increase if the elements increase and/or if the coefficient of the greater element increases. In terms of sets, if $pa + \bar{p}b > m$ and $a \geq b$, $+((p, 1], (a, 1], (b, 1]) \subseteq (m, 1]$. Conversely, if $a \leq b$, $+([0, p), (a, 1], (b, 1]) \subseteq (m, 1]$. Now, let $pa + \bar{p}b = m > r$. Then $pa + \bar{p}b > \frac{m+r}{2}$. If $a \geq b$, $+((p, 1], (a, 1], (b, 1]) \subseteq \left(\frac{m+r}{2}, 1\right] \subseteq (m, 1]$. If $a < b$, $+([0, p), (a, 1], (b, 1]) \subseteq \left(\frac{m+r}{2}, 1\right] \subseteq (m, 1]$. Therefore, every point in $+^{-1}((r, 1])$ has an open neighborhood, and $+^{-1}((r, 1])$ is an open set. This means that the mixing operation is continuous.

1.12 Lessons learned

This section summarizes insights from previous attempts that failed. We collect them here for reference.

Entropy and convex structure

Insight 1.217. *The entropy is not uniquely determined by the convex structure.*

When defining the entropy given a space of probability distributions, there is typically an underlying assumption that all "pure states" are equivalent. That is, they have the same entropy. For example, in a discrete classical space, the entropy is typically $-\sum_i p_i \log p_i$ which assumes the entropy is zero for all extreme points. This assumption is not tenable in general.

Example 1.218 (Mixed gas). Let's assume we have a mixture of two gases at a fixed temperature and volume. The state is therefore defined by the variables n_a and n_b that represent the number of molecules of the respective gases. The entropy of each configuration is $\log 1 = 0$. If we have a uniform distribution over $[0, N_a]$ for n_a and $[0, N_b]$ cases of n_b, the total number of cases is $(N_a+1)(N_b+1)$, and therefore the entropy of the joint state is $\log((N_a+1)(N_b+1))$. Now, suppose we cannot control the number of molecules for each gas, but only the total n. Each value of n will not correspond to the same variability of the elements of the ensemble. If we fix n number of molecules, in fact, there will be $n+1$ configurations possible, and therefore the entropy of each case should be $\log(n+1)$.

If we have the space of probability distributions over a single variable that represents the total number of molecules, then we would not be able to know whether these are all indistinguishable or divided into two distinct types. Therefore we would not know which is the correct physical entropy.

Note that in quantum mechanics there is a tacit assumption that all pure states have the same entropy. We could imagine, however, a theory where the pure states, the ones that we can ideally prepare and control, are not all at the same entropy. This would bake into the state space that not all evolutions are reversible. There is no reason to exclude this case, and it may turn out to be what we need for future theories.

Continuity of affine combinations

Insight 1.219. *Mixing cannot be a homeomorphism.*

To answer whether the embedding of the ensemble space in the topology of the vector space is continuous, we wanted to understand whether the inverse of the mixing is continuous. It turns out, it can't be in general because of the extreme points.

Proposition 1.220. *Let $\mathcal{E} = [0,1]$ and suppose that mixing is a homeomorphism, meaning $+_{p\mathsf{a}}(\mathsf{b}) = p\mathsf{a} + \bar{p}\mathsf{b}$ is a homeomorphism onto its image. Then $\mathcal{E}$ would have to have the discrete topology, which is not possible, since the topology must be second countable.*

Proof. Consider $+_{\frac{1}{2}0}([0,1]) = [0, \frac{1}{2}]$. Since $[0,1] = \mathcal{E}$ is an open set, then $[0, \frac{1}{2}]$ is an open set as $+_{\frac{1}{2}0}([0,1])$ is a homeomorphism. Similarly, $+_{\frac{1}{2}1}([0,1]) = [\frac{1}{2}, 1]$ is an open set. This means that $\{\frac{1}{2}\} = [0, \frac{1}{2}] \cap [\frac{1}{2}, 1]$ is an open set. This means that, for every $p \in [0,1]$, $+_{p0}(\{\frac{1}{2}\}) = \{\frac{1}{2}(1-p)\}$ is an open set. Similarly, for every $p \in [0,1]$, $+_{p1}(\{\frac{1}{2}\}) = \{\frac{1}{2}(1+p)\}$ is an

open set. This means that every singleton is open, and the topology is discrete. The topology, then, cannot be second countable and $\mathcal{E}$ cannot be an ensemble space. □

Note that, in the above example, $+_{\mathsf{ab}} : [0,1] \to \mathcal{E}$ would also fail to be a homeomorphism, though it is not clear how this is related to the failure of continuous embedding in a topological vector space (probably failure of scalar multiplication?).

The continuity of the mixing function means we can stretch open sets and still have open sets, but not, in general, shrink them. In retrospect, this makes sense as the bounds created by the boundaries of the convex set are of a different nature than the bounds of a finite precision measurement.

Convergence of points without convergence of fraction

Insight 1.221. *In a limit, the fraction does not necessarily converge to one.*

Since the fraction tells us how much an ensemble is part of another, an intuitive conjecture would be that if $\mathsf{a}_i \to \mathsf{a}$ then $\mathrm{frac}_{\mathsf{a}}(\mathsf{a}_i) \to 1$. That is, the fraction increases as we get closer to the limit. This does not work because one can make a limit over the extreme points.

Example 1.222 (Pure state convergence in a quantum space). Let ψ_i be a sequence of pure states and ψ a pure state such that $\psi_i \to \psi$ and $\psi_i \neq \psi$ for all i. For example, on a Bloch sphere, it would be a sequence of points on the surface that converges to a point on the surface. Then, since they are all pure states, we have $\mathrm{frac}_{\psi}(\psi_i) = 0$. This means $\psi_i \to \psi$ while $\mathrm{frac}_{\psi}(\psi_i) \to 0 \neq 1$.

Measures and topology

Insight 1.223. *Boundaries of open sets can have non-zero measure.*

An earlier attempt when trying to find a compatibility condition between measures and topology was to impose that boundaries of open sets have measure zero. The idea was that we cannot associate a non-zero measure to non-terminating conditions. Though the conceptual idea is sensible, the specific implementation does not work. We can find a counterexample on the real line.

Proposition 1.224. *Let μ be the Lebesgue measure on the real line $\mathbb{R}$. There exists an open set U such that $\mu(U) \neq \mu(\overline{U})$.*

Proof. Let C be the SmithVolterraCantor set. This is a closed set with no interior for which $\mu(C) = \frac{1}{2}$. Let $U = (0,1) \setminus C$. We have that U is an open set and $\overline{U} = [0,1]$. We have $\mu(\overline{U}) = \mu([0,1]) = 1$ while $\mu(U) = \mu((0,1)) - \mu(C) = \frac{1}{2}$. □

Subspaces from convex structure

Insight 1.225. *The convex structure, by itself, cannot define subspaces.*

Initially, we tried recovering the notion of subspaces purely from the convex structure. We did make some progress in the finite-dimensional case, but ultimately this does not work. We identified two problems.

The original definition was as follows.

Definition 1.226. *Let $\mathcal{E}$ be a convex space and $X \subseteq \mathcal{E}$ be a subset. We say that X is a **subspace** of $\mathcal{E}$ if it contains all the convex combinations and all the components of its elements. That is, for every $\mathsf{e}_1, \mathsf{e}_2, \mathsf{e}_3 \in \mathcal{E}$ and $\lambda \in (0,1)$ such that $p\mathsf{e}_1 + \bar{p}\mathsf{e}_2 = \mathsf{e}_3$ we have:*

- *$\mathsf{e}_1, \mathsf{e}_2 \in X$ implies $\mathsf{e}_3 \in X$*
- *$\mathsf{e}_3 \in X$ implies $\mathsf{e}_1, \mathsf{e}_2 \in X$.*

*The **convex span** of X is the smallest subspace containing X.*

Remark. As defined, the convex span of two elements will include all their possible mixtures (i.e. the segment that connects them), all possible decompositions (i.e. all lines that pass through them) plus, recursively, all other mixtures and decompositions that can be reached from those. Physically, the idea is that if we act on some ensembles, then we are also acting on all their components and mixtures. Therefore, the proper definition of subspace does not include just the mixtures, but all possible components and all their possible mixtures.

Example 1.227 (Classical discrete spaces). Let S be a set of n possible discrete states and let $\mathcal{E}$ be the space of probability distributions over the set S (i.e. $\mathcal{E}$ is an n-simplex and S its extreme points). A subspace X of $\mathcal{E}$ is a convex hull of a subset U of S. That is, a subspace of $\mathcal{E}$ is the space of probability distributions over a subset of the cases. Geometrically, it is one of the sides (possibly recursively) of the simplex.

To see this, first note that the convex hull X of any subset U of extreme points S is a subspace. In fact, it will contain all convex combinations of U, and any element can be decomposed in convex combinations of only U. Second, note that only convex hulls of a subset of extreme points can be a subspace. In fact, any element of $\mathcal{E}$ can be expressed as a non-trivial convex combination of a set of extreme points U. Therefore, if an element is present in a subspace X, then $U \subset X$, which means all elements of the convex hull of U are in X.

Example 1.228 (Finite-dimensional quantum spaces). Let $\mathcal{H}$ be an n-dimensional Hilbert space and let $\mathcal{E}$ be the space of density matrices (i.e. positive semi-definite self-adjoint operators with trace one). A subspace X of $\mathcal{E}$ is the space of density matrices of a subspace U of $\mathcal{H}$. That is, a subspace of $\mathcal{E}$ is the space of mixed states over a subspace of pure states.

To see this, first note that the space of density matrices X of a subspace U of $\mathcal{H}$ is a subspace of $\mathcal{E}$. In fact, X will contain all convex combinations of its elements. Moreover, any element $x \in X$ can be decomposed in a convex combination of pure states of only U. Therefore any convex decomposition of x has all its elements in X. Second, note that only the space of density matrices X of a subspace U of $\mathcal{H}$ is a subspace of $\mathcal{E}$. In fact, any element x of $\mathcal{E}$ can be expressed as a non-trivial convex combination of orthogonal pure states, its eigenstates. These elements will span a subspace U of $\mathcal{H}$. From those elements, we can construct an equal mixture which represents the maximally mixed states and, mathematically, is the identity operator I/m divided by the number of elements $m \leq n$ of U. The equal mixture of any orthogonal basis of U will also give the maximally mixed state. Therefore, given an element x, any subspace that contains x will also contain a basis of U, the maximally mixed state I/n, all possible basis of U, which means all the pure states, and finally all convex combinations of the pure states, which means all possible density matrices, all possible mixed states.

The definition works in these cases, which is why it looked promising, but not in general.

Insight 1.229. *Rate of convergence cannot be changed by a convex combination.*

Proposition 1.230. *Let $\mathcal{E}$ be the space of probability measures over $[0,1]$. Let* a *be the uniform distribution and let* b *be a distribution whose density goes to zero at the endpoints. Then* b *is a component of* a *but not vice-versa. Therefore the convex span of* a *is the whole $\mathcal{E}$ while the convex span of* b *is not the whole $\mathcal{E}$.*

Proof. Consider the two probability densities ρ_a and ρ_b associated with the ensembles. The density ρ_b has a supremum $\sup \rho_b$ that is greater than 1, which means $p = \frac{1}{\sup \rho_b}$ is smaller than one. The function $\rho_c = \frac{1}{\bar{p}}\rho_a - \frac{p}{\bar{p}}\rho_b$ will be a non-negative function that integrates to one. The same procedure could be applied for any probability density, which means every probability measure over $[0,1]$ is a component of the uniform distribution. Therefore the convex span of a is the whole $\mathcal{E}$.

Conversely, consider a possible convex combination $p\rho_a + \bar{p}\rho_c = \rho_b$. Since ρ_b converges to zero at the endpoints, for each p there will be a neighborhood of 0 for which $p\rho_a$ is greater than ρ_b. Therefore ρ_c cannot be a non-negative function. This means that a is not a component of b and the convex span of b is not the whole $\mathcal{E}$. □

This means that this notion of subspace does not simply return all the functions with the same support, but the function with the same support and a particular class of convergence on the boundaries, and possibly at interior points.

One key problem is that it cannot be generalized to the classical continuous case, as the rate of convergence cannot be changed by finite convex combinations. Even if that problem were fixed, there would be nothing to determine the dimensionality of U, and we could create convex maps that "stretch" the space. This led to the following:

Insight 1.231. *Dimension/count of states cannot be determined by the convex structure.*

Even if the previous problem is solved, the "size" of the subspace cannot be determined with a convex structure alone. In the finite-dimensional case this works, but in the infinite-dimensional case it does not.

Proposition 1.232. *Let $X = M_1([0,1])$ and $Y = M_1([0,2])$ be the spaces of probability measures on $[0,1]$ and $[0,2]$ respectively. The function $f : [0,1] \to [0,2]$ such that $f(x) = 2x$ is a diffeomorphism and induces a bijective continuous affine map between $M_1([0,1])$ and $M_1([0,2])$. However, the entropy is not conserved through the map.*

Proof. In terms of the probability measures, the map simply acts on the transformed sets. That is, $\mu_Y(A) = \mu_X(f^{-1}(A))$. This, for example, maps the uniform distribution over $[0,1]$ to the uniform distribution over $[0,2]$. Therefore A and B are isomorphic as convex spaces. However, the entropy over a uniform distribution changes with the range, therefore the entropy is not conserved through the map. □

This means that the convex structure and the topology are not enough to determine the size of the space. This means that, without information about the entropy, we are not going to be able to tell that we spread ensembles over double the distance.

Spectrum from subspaces

Insight 1.233. *The lattice of subspaces is not enough to recover the points.*

Originally, we intended to recover the spectrum of an operator (i.e. the possible values over which the distribution is defined) by taking the limit of subspaces. For example, for position a possible value x would be the limit of the sequence of subspaces of distributions with support $[x-\epsilon, x+\epsilon]$. Formally, the construction was similar to Stone's representation theorem for Boolean algebras, and the points were the ultrafilters. However, this does not work as the ultrafilters are "too fine." When reconstructing a real line, for example, the limit approaching from below and the limit approaching from above would be two separate points.

The core of the problem is that the lattice of subspaces recovers the Boolean algebra of the regular open sets. However, the lattice of the regular open sets is not enough to reconstruct the points of the space.

Proposition 1.234 (Regular open sets are not enough)**.** *Let X and Y be two topological spaces such that the respective lattices of regular open sets R_X and R_Y are isomorphic as Boolean algebras. It does not follow that X and Y are homeomorphic as topological spaces.*

Proof. Let $X = S^1$ be a circle with the standard topology, $Y = \mathbb{R}$ be the real line with the standard topology and $Z = \mathbb{R} \setminus \{0\}$ be the real line without the origin. Note that the circle is the one-point compactification of the real line, and therefore $Y = X \setminus \{\infty\}$. This forms a chain of embedding $Z \to Y \to X$.

Suppose U_X is a regular open set in X. Then $U_Y = U_X \setminus \{\infty\}$ is an open set in Y. We also have $\overline{U_Y} = \overline{U_X} \setminus \{\infty\}$ and $\mathrm{int}(\overline{U_Y}) = \mathrm{int}(\overline{U_X}) \setminus \{\infty\}$ where the closure and interior are taken in the respective spaces. Therefore U_Y is a regular open set. Also, let U_X and V_X be two distinct open sets of X. Then, since every open set that contains $\{\infty\}$ also contains an open neighborhood of $\{\infty\}$, the corresponding U_Y and V_Y will also be two distinct open sets. Then $\iota(U_X) = U_X \setminus \{\infty\}$ is an injection between the regular open sets of X and Y. Now let U_Y be a regular open set on Y. Let $U_X = \mathrm{int}(\overline{U_Y \cup \{\infty\}})$ where the closure and interior are taken in X. Then U_X is a regular open set of X because it is the interior of a closure. Also note that $\mathrm{int}(\overline{\iota(U_X) \cup \{\infty\}}) = \mathrm{int}(\overline{U_X \setminus \{\infty\} \cup \{\infty\}}) = \mathrm{int}(\overline{U_X}) = U_X$. Therefore the map ι is a bijection between the regular open sets of X and Y.

Note that regular open sets form a Boolean algebra where the complement is the exterior, the join is the interior of the closure of the union and the meet is the intersection. Note that if $U_X \subset V_X$, then $\iota(U_X) \subset \iota(V_X)$, therefore ι preserves the joins and the meets. Now let $V_X = \mathrm{ext}(U_X)$ and consider $\iota(V_X)$. This will contain all the exterior points of U_X minus $\{\infty\}$. But these are also the exterior points of $\iota(U_X)$. Therefore $\mathrm{ext}(\iota(U_X)) = \iota(\mathrm{ext}(U_X))$. The bijection is an isomorphism of Boolean algebras, and the corresponding algebras of regular open sets are isomorphic. The same reasoning applies when comparing Y to Z.

Now, note that $\{\infty\}$ is a closed set, but not a regular closed set. Therefore we are going to be able to construct a similar argument for the lattice of regular closed sets □

Since the construction in terms of subspaces recovered the algebra of regular open (or closed) sets, and wanted to recover the points as limits, the above proposition implies that this is not possible. We would not be able to know, for example, whether the points form a circle or a real line. Even if we knew that they form a real line, and knew what half-lines we have, we would not know how to stitch them back together, as we could do it with $\{0\}$ or with $\{\infty\}$. We would need to know, at least, which intervals correspond to a finite interval (e.g. a finite entropy), which requires more structure.

Open problems

This a collection of open problems related to the topics of this chapter. Here we summarize the status of some of them, which may include failed attempts and wrong results.

1.13 Open problem: Affine sets and affine hulls

Tags: Ensemble spaces, Convex spaces, Affine spaces

In the context of ensemble spaces, we have defined closures in terms of convex combinations. It may be useful to also define closures in terms of affine combinations. It needs to be understood whether this is useful, and how to make sure that these closures work for infinite dimensional spaces.

The following notes represent an early attempt. It needs to be understood whether this is at all needed and, in that case, revised and finalized. We used the term flat in this notes, but if may be better to first define an affine set as one closed under affine combinations and then the affine hull. This would be more consistent with the previous definitions and some of the literature.

In some cases, it may useful to define sets closed under affine combinations instead of convex combinations. That is, if $\mathsf{a} = p\mathsf{b} + \bar{p}\mathsf{c}$ and two of the ensembles are in the set, then the other one is in the set as well. This allows us to get all the ensembles in the same hyperplane. We call this a flat. In the triangle example, the only flats available are the single vertexes, each side and the whole triangle. For the Bloch ball, if we took two elements, we would get the whole segment that connects them and extends to the surface. But if we get three points, we get a circle, which is not a simplex. This affine closure, essentially, allows us to remove "ensembles from each other" as much as possible, so that we can get to their most distinct, most separate, form.

It still needs to be understood what the best definition for flat is, its relationship with the topological closure, its corresponding hull and so on.

Definition 1.235. *A **simplex** is a subset $U \subset \mathbb{R}^n$ that is the hull of finitely many affinely independent points. That is, $U = \text{hull}(\{x_i\}_{i=0}^n)$ such that $\{x_i - x_0\}_{i=1}^n$ are linearly independent.*

Definition 1.236. *A **flat** $A \subseteq \mathcal{E}$ is a closed convex subset that contains all lines between all elements. That is, for any $\mathsf{a}, \mathsf{b} \in A$, A also contains the line that contains a and b. Given a set $U \subseteq \mathcal{E}$, the **flat closure** of U is the smallest flat that contains U. A flat is **finite** if it can be generated by a finite number of elements. An n-flat is a flat that must be generated by a set with at least n elements.*

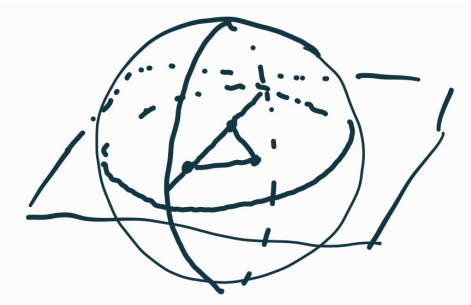


Figure 1.13: Flat vs convex set. The ball is the ensemble space. The triangle represents a convex subset. The disc containing the triangle is the flat closure of the triangle. The plane is the affine closure of the triangle.

Proposition 1.237. *A flat includes all possible affine combinations of elements of $\mathcal{E}$ that are contained in $\mathcal{E}$. The flat closure of $U \subseteq \mathcal{E}$ is the intersection of $\mathcal{E}$ with the affine closure of U in the embedding vector space.*

Proof. Let $A \subseteq \mathcal{E}$ be a flat and let V be the real vector space that embeds $\mathcal{E}$. A line between two elements $\mathsf{a}, \mathsf{b} \in A$ can be extended in V and can be written as $\mathsf{a} + x(\mathsf{b} - \mathsf{a})$ with $x \in \mathbb{R}$. Therefore any ensemble that can be written as an affine combination of two ensembles in A is an element of the flat. Recursively, this means that any ensemble that can be written as an affine combination of finitely many ensembles in A is also in the flat. Moreover, since the flat is a closed set, it will also include all its limits. Therefore every affine combination of A that is also an element of $\mathcal{E}$ is in A, which means A is the intersection of an affine subspace of the embedding vector space and the ensemble space.

Now let $U \subseteq \mathcal{E}$ be a set of ensembles. The affine closure of U will be an affine subspace of the embedding vector space. The flat closure will be the intersection of the affine closure with $\mathcal{E}$. Since the affine closure is the smallest closed affine subspace that contains U, its intersection with $\mathcal{E}$ will be the smallest flat that contains U, which is the flat closure of U. □

Conjecture 1.238. *An n-flat is a simplex if and only if its 3-flats are simplexes (i.e. triangles).*

Remark. In classical discrete ensemble spaces, any flat is a simplex. In classical continuous ensemble spaces, infinite dimensional flats will depend on what limits are allowed. In a quantum ensemble space, if we take three ensembles inside a Bloch ball, the corresponding flat will be a circle. If we take three orthogonal pure states, however, the corresponding flat will be a simplex.

1.14 Open problem: Classicality as reducibility

Tags: Ensemble spaces, Convex spaces

We still need a solid characterization of the ensemble space for classical theories. Effectively, we are looking for a characterization of a convex space to be a symplex without extreme points, such that the extreme points can be understood as limits, and not elements of the convex space.

Part of the problem is isolating the different parts that may lead to different features that one may consider classical. For example, finite discrete classical spaces are simplexes such that all extreme points are orthogonal to each other and have zero entropy. One may have a simplex where all extreme points are orthogonal, though the extreme points have different entropy. All features of a classical theory are there, expect that the entropy function has to be corrected to $S(\mathsf{e}) = I(p_i) + \sum_i p_i S(x_i)$. A weaker case is when the ensemble space may be a simplex, but the extreme points are not orthogonal.

The definition will need to work for the classical continuous case. Given that classical phase space is not, in general compact, the space of probability distribution is not a Choquet simplex. Moreover, even if we restrict to ourselves to the compact case, the Dirac measures cannot be in the ensemble space, as their entropy is minus infinity.

Here, we focus on the intermediate property: the ensemble can be orthogonally decomposed, but the entropy of the extreme points is not necessarily uniform.

Definition 1.239. *An ensemble space $\mathcal{E}$ is **reducible** if every decomposable ensemble can be decomposed into ensembles with no common components and ensembles that have no common component are mutually exclusive. That is, it is separately decomposable and separateness implies orthogonality (i.e. $\mathsf{e}_1 \Subset \mathsf{e}_2$ implies $\mathsf{e}_1 \perp \mathsf{e}_2$).*

Justification. The idea of reducibility is that we can take ensembles and decompose them into parts that are mutually exclusive. This is the basic expectation in the classical case.

If we assume reducibility, then, if an ensemble is a mixture of other ensembles, we must be able to express it as a mixture of separate ensembles. This justifies separate decomposability.

Given that $\mathsf{e}_1 \perp \mathsf{e}_2$ implies $\mathsf{e}_1 \Subset \mathsf{e}_2$, reductibility add the opposite implication. Therefore it excludes the case where two ensembles are orthogonal but not separate. This case describes two ensembles that have elements in common, but there is no ensemble corresponding to those common elements. That is, we cannot refine the ensembles into three separate ones: one with only elements of the first, one with only elements of the second and one with elements of both. In other words, we cannot reduce the coarser description of the system into finer separate descriptions. In the excluded case, then, the coarser description is irreducible into finer ensembles. This justifies that, under reducibility, separate ensemble must be orthogonal. □

Proposition 1.240. *Continuous and discrete classical ensemble spaces are reducible.*

Proof. As we saw before, classical ensemble spaces are separately decomposable. We also saw that both separateness and orthogonal both correspond to disjoint support. Therefore separateness implies orthogonality. □

What we need to prove is that every reducible ensemble space can be represented by a subspace of probability measures. The strategy would be to prove the following series of conjectures. The lattice of $\perp$-subspaces $\mathfrak{L}^\perp$ forms a Boolean algebra. Using Stone's representation theorem for Boolean algebras, there is a set X and $\mathfrak{L}^\perp$ correspond to the clopen sets, and orthogonal subspaces correspond to disjoint sets. Set functions over clopen sets can be extended to the Borel algebra of X. Since the state capacity is additive over orthogonal sets, it is an additive set function when defined to the Borel algebra of X and, therefore, it is a measure. Similarly, one finds that the fraction capacity becomes additive, and, given its other properties, it is a probability measure over X. Additionally, since every ensemble has a finite

entropy and the state capacity gives us the supremum of the exponential of the entropy, the fraction capacity for a set of measure zero (i.e. minus infinity entropy) must be zero (i.e. no component can exist).

Conjecture 1.241. *Every reducible ensemble space can be represented by a space of probability measures. That is, given a reducible ensemble space $\mathcal{E}$ there is a topological space X such that each $\mathsf{e} \in \mathcal{E}$ is represented by a unique probability measure p_{e} over X that is absolutely continuous with a measure μ over X.*

1.15 Open problem: Classical contexts

Tags: Ensemble spaces, Convex spaces

Even if an ensemble space is not classical, we want to be able to characterize subsets of ensembles spaces that look classical. In quantum mechanics, this would recover the contexts in which classical probability can be defined

There is again the difference of just considering the convex structure (i.e. they form a simplex) or also add the entropic structure (i.e. separate ensembles are orthogonal). In the attempt below, we chose to only use the convex structure. These are still rough ideas that may need to be refined.

The basic idea is that the hull of a set of orthogonal states in quantum mechanics forms a simplex, so we can understand projection measurements as moving the system to one of those states. Ideally, a classical context should be a subset of an ensemble space that is a classical ensemble space. An additional problem is the multiplicity of extreme points for the same measurement outcome. In that case, the set of post-measurement ensembles are not a simplex. It may be that the correct definition for a classical context would be the domain of a map to simplex that preserves orthogonality.

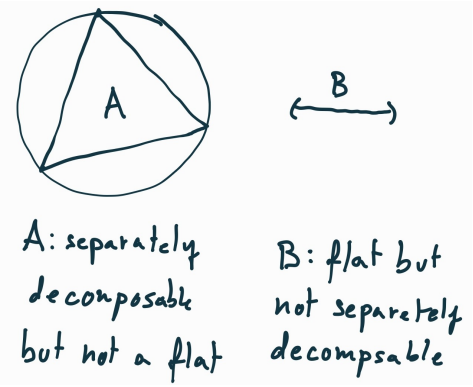
Sets decomposability

Since contexts set can be a subset of, for example, a quantum ensemble space and this property must work recursively, the first step is to restring the notion of decomposability to sets.

Definition 1.242. *Let C be a convex set. The set is **separately/orthogonally decomposable** if every element of C that is decomposable in C is also separately/orthogonally decomposable in C. The set is **monodecomposable** if every element in C that is decomposable in C is monodecomposable in C. The set is **separately/orthogonally monodecomposable** if it is both separately/orthogonally decomposable and monodecomposable.*

Definition 1.243. *Let C be a convex set. We say that **mixtures preserve separateness in** C if $\mathsf{e} \pitchfork \mathsf{a}$ and $\mathsf{e} \pitchfork \mathsf{b}$ implies $\mathsf{e} \pitchfork p\mathsf{a} + \bar{p}\mathsf{b}$ for all $p \in [0,1]$.*

Proposition 1.244. *A convex set C is monodecomposable in C if and only if mixtures preserve separateness in C.*



Remark. A convex set that is separately decomposable is not necessarily a flat (e.g. triangle within a sphere). A flat is not necessarily separately decomposable (e.g. open segment as a whole is a flat, but is not separately decomposable).

We can now define a classical probability context as a flat that does not allow multiple decompositions. The definition below may still need to be refined. In particular, we may want need to include the requirement of orthogonality during the decomposition for an actual classical context as classical probability implicitly requires that the cases are mutually exclusive. We would need a different name when the requirement is only on the convex structure, like below.

Definition 1.245. *Let $\mathcal{E}$ be an ensemble space. A **classical probability context** is a flat $C \subseteq \mathcal{E}$ where each decomposable element in C is separately decomposable in C but not separately multidecomposable in C.*

The idea should be that we can use the lack of multidecomposability to split classical contexts recursively.

Proposition 1.246. *A flat $C \subseteq \mathcal{E}$ is a classical probability context if and only if every decomposable element is separately decomposable and mixtures preserve separateness in C. That is, if $\mathsf{e} \top \mathsf{a}$ and $\mathsf{e} \top \mathsf{b}$, then $\mathsf{e} \top p\mathsf{a} + \bar{p}\mathsf{b}$ in C for all $\mathsf{e}, \mathsf{a}, \mathsf{b} \in C$ and $p \in [0,1]$.*

Proof. Suppose mixtures do not preserve separateness. Then we can find $\mathsf{e}, \mathsf{a}, \mathsf{b}, \mathsf{c} \in \mathcal{E}$ such that $\mathsf{e} \top \mathsf{a}$, $\mathsf{e} \top \mathsf{b}$ and $\mathsf{e} \perp \mathsf{c} = p\mathsf{a} + \bar{p}\mathsf{b}$ for some $p \in (0,1)$. Since $\mathsf{e} \perp \mathsf{c}$, we can find $\mathsf{d}, \mathsf{f}, \mathsf{g}$ such that $\mathsf{c} = \lambda\mathsf{d} + \bar{\lambda}\mathsf{f}$ and $\mathsf{e} = \mu\mathsf{d} + \bar{\mu}\mathsf{g}$. Since separateness extends to all mixtures (1.24) and $\mathsf{a} \top \mathsf{e}$, $\mathsf{a} \top \mathsf{d}$ and, similarly, $\mathsf{b} \top \mathsf{d}$, which means that c is separately multidecomposable.

Now suppose mixture do preserve separateness, and let $\mathsf{e} = p\mathsf{a}_1 + \bar{p}\mathsf{a}_2 = \lambda\mathsf{b}_1 + \bar{\lambda}\mathsf{b}_2$. Since a_1 is a component of e, then $\mathsf{e} \perp \mathsf{a}_1$. Since e is a mixture of b_1 and b_2 and mixtures preserve separateness, then either $\mathsf{a}_1 \perp \mathsf{b}_1$ or $\mathsf{a}_1 \perp \mathsf{b}_2$. Similarly, $\mathsf{a}_2 \perp \mathsf{b}_1$ or $\mathsf{a}_2 \perp \mathsf{b}_2$. Therefore e is not separately multidecomposable. □

Conjecture 1.247. *A convex subset $U \subseteq \mathcal{E}$ is a classical probability context if and only if all finite flats are simplexes.*

Proposition 1.248. *Let $C \subseteq \mathcal{E}$ be a classical probability context. Let $U \subseteq C$ be a set of ensembles. Then $A = U^{\top} = \{\mathsf{a} \in C \mid \forall \mathsf{e} \in U, \mathsf{a} \top \mathsf{e}\}$ and $B = (U^{\top})^{\top} = \{\mathsf{b} \in C \mid \forall \mathsf{a} \in A, \mathsf{b} \top \mathsf{a}\}$ are two probability contexts such that $C = \mathrm{hull}(A \cup B)$.*

Proof. First we show that C contains only three types of ensembles: those that are limits of convex mixtures of A, those that are limits of convex mixtures of B and those that are limits of convex mixtures of both A and B. Any ensemble in C is one of these three types. For $\mathsf{c} \in C$ to not be a mixture of A or B, then c cannot be in either A or B. This means that it must be separate from both A and B. But B contains all the elements that are separate from A, which is a contradiction. Since separate multidecomposition is forbidden, an ensemble e cannot be written both as a convex combination of A and as a convex combination with an element of B. This would yield two decompositions in which one component, the one chosen from B, is separate from all the components of the other. This means that an element of C is either a mixture of A, a mixture of B or a mixture of both A and B.

Now we show that both A and B are convex sets. Since a mixture of A can only be expressed as a mixture of components of A, it is separate from all elements of B. Therefore A contains all its mixtures. With the same logic, B will contain all its mixtures. The argument works for infinite convex combinations as well. If $\mathsf{e} = \sum_{i=1}^{\infty} p_i \mathsf{a}_i$ with $p_i \in [0,1]$ and $\sum_{i=1}^{\infty} p_i = 1$, it can be understood as the limit of the series $\frac{p_1}{P_n}\mathsf{a}_1 + \frac{\bar{p}_1}{P_n}\sum_{i=2}^{n} p_i \mathsf{a}_i$ where $P_n = \sum_{i=1}^{n} p_i$. Therefore

$$\begin{aligned}\mathsf{e} &= \sum_{i=1}^{\infty} p_i \mathsf{a}_i = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{p_i}{P_n}\mathsf{a}_i = \lim_{n\to\infty} \frac{p_1}{P_n}\mathsf{a}_1 + \lim_{n\to\infty} \frac{\bar{p}_1}{P_n}\sum_{i=2}^{n} \frac{p_i}{\bar{p}_1}\mathsf{a}_i = p_1\mathsf{a}_1 + \bar{p}_1 \sum_{i=2}^{\infty} \frac{p_i}{\bar{p}_1}\mathsf{a}_i \\ &= p_1\mathsf{a}_1 + \bar{p}_1\hat{\mathsf{a}}_1\end{aligned} \tag{1.249}$$

where $\hat{\mathsf{a}}_1 = \sum_{i=2}^{\infty} \frac{p_i}{\bar{p}_1}\mathsf{a}_i$. Since e and a_1 are elements of the ensemble space, the series converges to $\hat{\mathsf{a}}_1$, which is in the hull of A. It will also be an element of A because multidecompositions are forbidden.

To see that C is the hull of A and B, note that all the elements in C that are not already in A or B are the mixtures of A and B. These are exactly added when taking the hull of $A \cup B$.

Now we show that A and B are classical probability contexts. First we have to show that they are flats. Let L be the line that connects two elements $\mathsf{a}_1, \mathsf{a}_2 \in A$. Take $\mathsf{a}_3 \in L$. If it is a mixture of a_1 and a_2 then it is an element of A. If a_1 is a mixture of a_2 and a_3, since a_1 cannot have a common component with B, and a_3 is a component of a_1, a_3 cannot have components in B as well. Therefore a_3 must be a mixture of elements of A. Similarly if a_2 is a mixture of a_1 and a_3. Therefore A is a flat. Similarly, B is a flat.

Now we show that A, and by symmetry B, is a classical probability context. We have seen that A is a flat. If an element of A is decomposable in A it is also decomposable in C and is therefore separately decomposable in C. Because multidecomposability in C is not allowed, it must be separately decomposable into elements of A. Lastly, if multidecomposability were allowed in A, it would also be allowed in C. Therefore it is not allowed in A. This means that A is a classical probability context. □

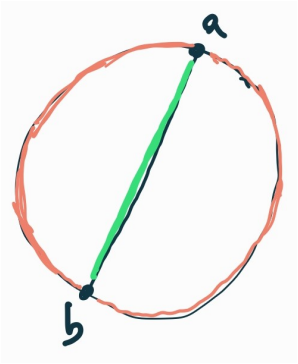
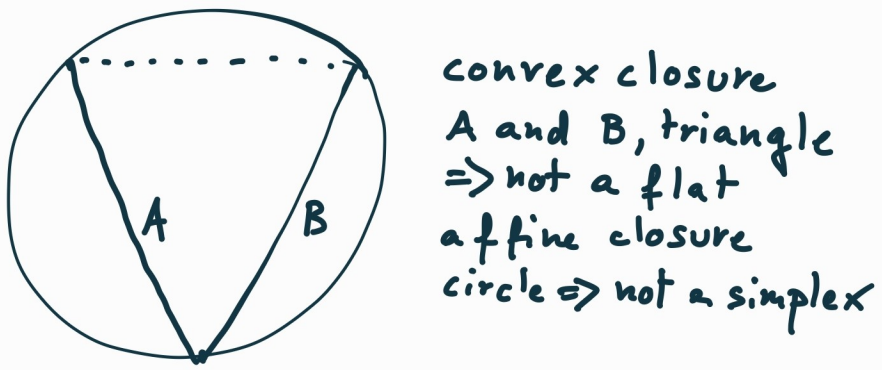


Figure 1.14: TODO: make the final picture not go through the center, change the label for the points to e_1 and e_2

Remark. Note that if multiple separate decompositions are not ruled out, the sub-contexts will not include all convex combinations. Take a disk as a convex space. Suppose U is made of two points e_1 and e_2. All other points on the surface are separate from both and therefore belong to A. Now consider the convex combinations of e_1 and e_2. These are not separate from U but they are also not separate from all the other elements on the surface. Therefore they are neither in A nor B. Moreover, any other point in the interior can be seen as a convex combination of e_1 and another element of the surface that is not e_2. Therefore no point in the interior is in either A or B. Thus A and B are not necessarily convex sets if C is not a classical probability context.



Remark. While a classical probability context can be decomposed into classical probability contexts whose closure is the original context, the converse is not true. That is, given two classical probability contexts, their convex or flat closure is not in general a classical probability context. Take a disk (the ensemble space) and two lines connecting three points (the two probability contexts). The convex closure is the triangle, which is not a flat as it misses some affine combinations. The flat closure is the disk, which is not a probability context.

Conjecture 1.250. *Let $A, B \subseteq \mathcal{E}$ be two classical probability contexts. Then their flat closure and convex closure coincide if and only if these closures are a classical probability context.*

Conjecture 1.251. *The defining property of a classical probability context is exactly that property that allows A and B so constructed to be convex sets whose closure is C.*

Proposition 1.252. *Let $C \subseteq \mathcal{E}$ be a classical probability context. The lattice of π-subspaces $\mathfrak{L}$ is distributive and is therefore a Boolean algebra.*

Proof. Let $X, Y, Z \in \mathfrak{L}$ be three π-subspaces. Let $\mathsf{e} \in X \wedge (Y \vee Z)$, then $\mathsf{e} \in X$ and $\mathsf{e} \in Y \vee Z$. This also means that all components of e are in both X and $Y \vee Z$ since the components of e are not separate from e and, given the absence of multi-decomposability, are separate from all elements that are separate from e. Since $Y \vee Z = \text{hull}(Y \cup Z)$, we can decompose e into components of Y and Z. But these must be in X, and therefore $\mathsf{e} \in \text{hull}((X \wedge Y) \cup (X \wedge Z)) = (X \wedge Y) \vee (X \wedge Z))$. Therefore $X \wedge (Y \vee Z) \subseteq (X \wedge Y) \vee (X \wedge Z)$. Now let $\mathsf{e} \in (X \wedge Y) \vee (X \wedge Z)$, then $\mathsf{e} \in \text{hull}((X \wedge Y) \cup (X \wedge Z))$. Therefore e is the mixture of ensembles that are in X, which means that $\mathsf{e} \in X$. Since e is the mixture of elements that are also in Y and Z, $\mathsf{e} \in \text{hull}(Y \cup Z) = Y \vee Z$. This means $\mathsf{e} \in X \wedge (Y \vee Z)$. Therefore $(X \wedge Y) \vee (X \wedge Z) \subseteq X \wedge (Y \vee Z)$. Which means $X \wedge (Y \vee Z) = (X \wedge Y) \vee (X \wedge Z)$.

Now let $\mathsf{e} \in X \vee (Y \wedge Z)$, then $\mathsf{e} \in \text{hull}(X \cup (Y \wedge Z))$. That is, e is a mixture of elements of X and of elements that are in both Y and Z. Which means $\mathsf{e} \in \text{hull}(X \cup Y)$ and $\mathsf{e} \in \text{hull}(X \cup Z)$. Therefore $\mathsf{e} \in \text{hull}(X \cup Y) \cap \text{hull}(X \cup Z) = (X \vee Y) \wedge (X \vee Z)$, which means $X \vee (Y \wedge Z) \subseteq (X \vee Z) \vee (Y \wedge Z)$. Conversely, let $\mathsf{e} \in (X \vee Y) \wedge (X \vee Z) = \text{hull}(X \cup Y) \cap \text{hull}(X \cup Z)$. Then e can be expressed as a mixture of elements of X and Y and also of elements of X and Z. Now consider the elements of Y in the first decomposition. Since multiple decomposition is not allowed, they cannot be separate from both X and Z, end therefore must be in their hull. Similarly, the elements of Z in the second decomposition must be in the hull of X and Y. Therefore $\mathsf{e} \in \text{hull}(X \cup (Y \wedge Z)) = X \vee (Y \wedge Z)$. Therefore $(X \vee Y) \wedge (X \vee Z) \subseteq X \vee (Y \wedge Z)$. Which means $X \vee (Y \wedge Z) = (X \vee Z) \vee (Y \wedge Z)$.

The lattice $\mathfrak{L}$ is therefore a distributive orthocomplemented lattice, which means it is a Boolean algebra. □

Proposition 1.253. *Let $C \subseteq \mathcal{E}$ be a classical probability context. Let $\mathfrak{L}$ be the lattice of π-subspaces. Then the fraction capacity is an additive set function on the lattice. That is, for all $\mathsf{e} \in C$ and $A, B \in \mathfrak{L}$ such that $A \cap B = \varnothing$, $\text{fcap}_{\mathsf{e}}(A \vee B) = \text{fcap}_{\mathsf{e}}(A) + \text{fcap}_{\mathsf{e}}(B)$.*

Proof. Let $A, B \in \mathfrak{L}$ such that $A \cap B = \varnothing$. Then A is separate from B. Therefore $A \vee B$ is a classical probability context that is the convex closure of two separate probability contexts. Therefore, as we saw in a previous proof, $A \vee B$ consists of convex combinations of A, which are all in A, of convex combinations of B, which are all in B, and of convex combinations of both, which are in neither A nor B.

We want to show that $\text{fcap}_{\mathsf{e}}(A \vee B) = \text{fcap}_{\mathsf{e}}(A) + \text{fcap}_{\mathsf{e}}(B)$ if $\mathsf{e} \in A \vee B$. Let $\mathsf{e} \in A \vee B$ be an ensemble that is the mixture of elements of A. Then $\mathsf{e} \in A$ and it has no components in B. Therefore $\text{fcap}_{\mathsf{e}}(A \vee B) = 1$ and $\text{fcap}_{\mathsf{e}}(A) = 1$ while $\text{fcap}_{\mathsf{e}}(B) = 0$. This means $\text{fcap}_{\mathsf{e}}(A \vee B) = 1 = 1 + 0 = \text{fcap}_{\mathsf{e}}(A) + \text{fcap}_{\mathsf{e}}(B)$. If e is a mixture of elements of B, we get the same conclusion. The last case is when e is a mixture of elements of A and B. That is, $\mathsf{e} = \sum_i p_i \mathsf{a}_i + \sum_j \lambda_j \mathsf{b}_j$ with $\mathsf{a}_i \in A$, $\mathsf{b}_j \in B$, $p_i, \lambda_j \in [0,1]$ and $\sum_i p_i + \sum_j \lambda_j = 1$. By definition of fraction capacity, $\sum_i p_i \leq \text{fcap}_{\mathsf{e}}(A)$. Suppose $\sum_i p_i < \text{fcap}_{\mathsf{e}}(A)$. Then there is $\hat{\mathsf{a}} \in A$ such that $\mathsf{e} = \sum_i p_i \mathsf{a}_i + p\hat{\mathsf{a}} + \lambda \mathsf{c}$ for some $p, \lambda \in [0,1]$ and $\mathsf{c} \in A \vee B$. But then e would be separately multidecomposable, which is a contradiction since it is an element of a classical probability context. Therefore $\sum_i p_i = \text{fcap}_{\mathsf{e}}(A)$. Similarly, we find that $\sum_j \lambda_j = \text{fcap}_{\mathsf{e}}(B)$. Since $\sum_i p_i + \sum_j \lambda_j = 1$, $\text{fcap}_{\mathsf{e}}(A) + \text{fcap}_{\mathsf{e}}(B) = 1 = \text{fcap}_{\mathsf{e}}(A \vee B)$ for all $\mathsf{e} \in A \vee B$.

Now let $\mathsf{e} \in C$. We have $(A \vee B)^{\mathbb{T}} \in \mathfrak{L}$ and $(A \vee B) \cap (A \vee B)^{\mathbb{T}} = \varnothing$. Therefore $\mathrm{fcap}_{\mathsf{e}}(C) = \mathrm{fcap}_{\mathsf{e}}(A \vee B) + \mathrm{fcap}_{\mathsf{e}}((A \vee B)^{\mathbb{T}})$. Note that e, in general, is a convex combination of elements of $A \vee B$ and $(A \vee B)^{\mathbb{T}}$. Components in $A \vee B$ are convex combinations of elements of A and B, which also disjoint. Therefore e is a convex combination of elements of A, B and $(A \vee B)^{\mathbb{T}}$, which are pairwise disjoint. As before, since multidecomposability is forbidden in a classical probability context, the sum of the coefficients for each part will have to match the fraction capacity of its subcontext. Therefore $\mathrm{fcap}_{\mathsf{e}}(A) + \mathrm{fcap}_{\mathsf{e}}(B) + \mathrm{fcap}_{\mathsf{e}}((A \vee B)^{\mathbb{T}}) = \mathrm{fcap}_{\mathsf{e}}(C) = \mathrm{fcap}_{\mathsf{e}}(A \vee B) + \mathrm{fcap}_{\mathsf{e}}((A \vee B)^{\mathbb{T}})$. Thus $\mathrm{fcap}_{\mathsf{e}}(A \vee B) = \mathrm{fcap}_{\mathsf{e}}(A) + \mathrm{fcap}_{\mathsf{e}}(B)$ for all $\mathsf{e} \in C$. □

Probability measures from classical probability contexts

The next step is to show that each ensemble in a classical probability context can be understood as a probability measure and vice-versa. The issue is how to recover the set of points, the spectrum $\sigma(C)$ of the context, and its topology and σ-algebra $\Sigma_{\sigma(C)}$. This is still an open problem. The Stone's representation theorem can be used as a fallback, though it does not recover the correct points.

Definition 1.254. *Let C be a context. An **spectral element** of C is a non-empty collection of subspaces $c \subset C$ such that*

1. *if $\{X_i\}_{i \in I} \in s$ then $\bigcap_{i \in I} X_i \in s$*
2. *if $X \in s$ and $Y \in \mathfrak{L}$ such that $X \subseteq Y$, then $Y \in s$*
3. *if $X \in s$ and $Y \in \mathfrak{L}$ such that $Y \subset X$ and $Y \neq \varnothing$, then there exists $Z \in s$ such that $Z \subset X$.*

*The set of all elements of the context is called the **spectrum** of the context and is noted $\sigma(C)$.*

Remark. The above should correspond to an ultrafilter.

Definition 1.255. *Let C be a context and $X \in C$ be a subspace. A **spectral element** of X is a spectral element $c \in \sigma(C)$ such that $X \in c$. The **spectral set** of X is the collection of all its spectral elements. The standard topology of the spectrum $\sigma(C)$ is the one generated by the spectral sets of all subspaces $X \in C$.*

Definition 1.256. *Given a poset $(P, \leq)$ downward refined filter is a proper filter F such that if $x \in F$ and there exists a $y \in P$ such that $y < x$, then there exists a $z \in F$ such that $z < x$.*

Definition 1.257. *Given a classical probability context $C \subseteq \mathcal{E}$, the **spectrum** $\sigma(C)$ is the set of equivalence classes of sequences of ensembles that eventually become separate from each other. More rigorously, given a probability context, the spectrum is the collection of the downward refined filters of the lattice of $\mathbb{T}$-subspaces $\mathfrak{L}^{\mathbb{T}}(C)$. The spectrum of each $\mathbb{T}$-subspace is given by $\sigma(A) \mapsto \{x \in \sigma(C) \,|\, A \in x\}$. The standard topology of the spectrum is the one generated by the spectra of all $\mathbb{T}$-subspaces (i.e. $\sigma(\mathfrak{L}^{\mathbb{T}}(C))$).*

Proposition 1.258. *Let $C \subseteq \mathcal{E}$ be a classical probability context and $\sigma(C)$ its spectrum. Then the following are true:*

1. *$A \subseteq B$ if and only if $\sigma(A) \subseteq \sigma(B)$*
2. *$\mathrm{ext}(\sigma(A)) = \sigma(A^{\mathbb{T}})$*
3. *$\partial\sigma(A) = \overline{\{x \in \sigma(C) \,|\, A \notin x, A^{\mathbb{T}} \notin x\}}$*
4. *$\sigma(A)^{\mathsf{C}} = \overline{\sigma(A^{\mathbb{T}})}$*

5. $\sigma(\bigwedge_{i\in I} A_i) = \mathrm{int}\left(\bigcap_{i\in I} \sigma(A_i)\right)$
6. $\sigma(\bigvee_{i\in I} A_i) = \mathrm{int}\left(\overline{\bigcup_{i\in I} \sigma(A_i)}\right)$
7. *if* U *open, then* $U = \bigcup_{i\in I} \sigma(A_i)$ *for some family of* $A_i \in \mathfrak{L}^{\Vdash}(C)$

Proof. For 1, since all $x \in \sigma(C)$ are upward closed, $A \in x$ means $B \in x$ as well. Therefore if $x \in \sigma(A)$ then $x \in \sigma(B)$, which means $\sigma(A) \subseteq \sigma(B)$. Conversely, if $\sigma(A) \subseteq \sigma(B)$ then all downward sets of $\mathfrak{L}^{\Vdash}(C)$ that contain A must also contain B, which means $B \supseteq A$.

For 2 and 3, note that $\sigma(A) = \{x \in \sigma(C) \mid A \in x\}$ and therefore $\sigma(A)^{\complement} = \{x \in \sigma(C) \mid A \notin x\}$. Since $\sigma(A)$ is an open set, $\sigma(A)^{\complement} = \mathrm{ext}(\sigma(A)) \cup \partial\sigma(A)$. Now consider $\sigma(A^{\Vdash})$. This is an open set and it is disjoint from $\sigma(A)$. In fact, if $x \in \sigma(A) \cap \sigma(A^{\Vdash})$ then $A, A^{\Vdash} \in x$. But this would mean that $\varnothing \in x$, which can't be because x is a proper filter and cannot contain $\varnothing$. Since $A^{\Vdash}$ is the largest $\Vdash$-subspace that is separate from A, there is no set in $\sigma(\mathfrak{L}^{\Vdash}(C))$ that is larger than $A^{\Vdash}$ and still disjoint from A. Therefore $\sigma(A^{\Vdash}) = \mathrm{ext}(\sigma(A))$. We have $\partial\sigma(A) = \sigma(C) \setminus (\mathrm{int}(\sigma(A)) \cup \mathrm{ext}(\sigma(A))) = \{x \in \sigma(C) \mid A \notin x, A^{\Vdash} \notin x\}$.

For 4, since $\sigma(A)$ is an open set, the complement is the closure of the exterior. Therefore, given 2, $\sigma(A)^{\complement} = \overline{\sigma(A^{\Vdash})}$.

For 5, consider $\bigwedge_{i\in I} A_i$. This is the largest $\Vdash$-subspace that is contained by all A_i. Then, given 1, $\sigma\left(\bigwedge_{i\in I} A_i\right)$ must be the largest open set that is contained by all $\sigma(A_i)$. This corresponds to $\mathrm{int}\left(\bigcap_{i\in I} \sigma(A_i)\right)$.

For 6, we have

$$\begin{aligned}
\sigma\Big(\bigvee_{i\in I} A_i\Big) &= \mathrm{ext}\Big(\mathrm{ext}\Big(\sigma\Big(\bigvee_{i\in I} A_i\Big)\Big)\Big) = \mathrm{ext}\Big(\sigma\Big(\Big(\bigvee_{i\in I} A_i\Big)^{\Vdash}\Big)\Big) = \mathrm{ext}\Big(\sigma\Big(\bigwedge_{i\in I} A_i^{\Vdash}\Big)\Big) \\
&= \mathrm{ext}\Big(\mathrm{int}\Big(\bigcap_{i\in I} \sigma(A_i^{\Vdash})\Big)\Big) = \mathrm{ext}\Big(\bigcap_{i\in I} \sigma(A_i^{\Vdash})\Big) = \\
&= \mathrm{ext}\left(\bigcap_{i\in I} \overline{\sigma(A_i)}^{\complement}\right) = \mathrm{ext}\left(\left(\bigcup_{i\in I} \overline{\sigma(A_i)}\right)^{\complement}\right) \\
&= \mathrm{int}\left(\bigcup_{i\in I} \overline{\sigma(A_i)}\right)
\end{aligned}$$

For the last step, note that the union of the closure is not necessarily the closure of the union, but they have the same interior.

For 7, note that an open set U is generated from $\sigma(\mathfrak{L}^{\Vdash}(C))$ through finite intersection and arbitrary union. Note that the finite intersection of open sets is an open set, so we have $\bigcap_{i\in I} \sigma(A_i) = \mathrm{int}\left(\bigcap_{i\in I} \sigma(A_i)\right) = \sigma(\bigwedge_{i\in I} A_i)$. Therefore the finite intersection corresponds to a $\Vdash$-subspace. This means that we can can generate all open sets with arbitrary unions of sets from $\sigma(\mathfrak{L}^{\Vdash}(C))$. □

Corollary 1.259. *For every* $A \in \sigma(\mathfrak{L}^{\Vdash}(C))$, $\sigma(A)$ *is a regular open set. The topology of a spectrum is semiregular.*

Proof. Since the lattice is orthocomplemented, we have $\sigma(A) = \sigma((A^{\Vdash})^{\Vdash}) = \mathrm{ext}(\sigma(A^{\Vdash})) = \mathrm{ext}(\mathrm{ext}(A))$. Therefore $\sigma(A)$ is a regular open set.

The topology is generated by regular open sets, and is therefore semiregular. □

Conjecture 1.260. *Given a point* $x \in \sigma(C)$ *and an open neighborhood* U *of* x*, we can find a closed neighborhood* A *that is a subset of* U*. The topology of a spectrum is regular.*

Proof. Since U is an open set, it is the union of a family of sets $\sigma(A_i)$. Since U is a neighborhood of x, $x \in \sigma(A)$ with $A = A_i$ for some i. Now consider $y \in \partial\sigma(A)$. Since both x and y are two distinct downward refined filters, there will be a π-subspace B_y such that $B_y \in x$ but $B_y \notin y$. That is, $x \in \sigma(B_y) \subseteq \sigma(A)$ and $y \notin \sigma(B_y)$. Now consider $D = \bigcap_{y \in \partial\sigma(A)} B_y$. Since $\mathfrak{L}^{\pi}(C)$ is an intersection structure, $D \in \mathfrak{L}^{\pi}(C)$. We also have $D \in x$ and therefore $D \neq \varnothing$. Therefore $\sigma(D)$ is an open set that is a subset of $\sigma(A)$ where we removed an open neighborhood around its boundary. This means that $\partial\sigma(D) \subseteq \sigma(A)$. This means that $\overline{\sigma(D)} \subseteq \sigma(A)$. Therefore we found a closed neighborhood of x that is a subset of U.

The above means that the closed neighborhoods of x form a local base for x, which is another way to characterize regular topological spaces. □

Remark. Note how the infinite joins/meets on the lattice of the π-subspaces do not correspond to the infinite operations on the lattice of the topology or of the Borel sets. For example, suppose we are taking the lattice of probability measures defined over the interval $[0, 1]$ that are absolutely continuous with respect to the Lebesgue measure. Each π-subspace will correspond to a set of measures whose support lives in a particular region. A singleton will not correspond to any subspace as it cannot support an absolutely continuous measure. Therefore, if we take all the π-subspaces of all measures that have $1/2$ within the support, the intersection of all those subspaces will be the empty set, which is a π-subspace. However, the intersection of the sets representing their support is the singleton $\{1/2\}$, which is not the empty set. However, its interior is the empty set.

The fact that a single probability measure is an equivalence class of probability densities is related to this distinction. Consider a uniform distribution over $[0, 1]$. The related probability density can be understood as a constant between those values and zero everywhere else. However, the same constant over $[0, 1/2) \cup (1/2, 1]$ will also work. That is, the same element of the ensemble space can be represented in different ways as a function over the spectrum.

Conjecture 1.261. *Let $C \subseteq \mathcal{E}$ be a probability context and $\mathsf{e} \in C$ be an ensemble in the context. The set function $p_{\mathsf{e}} : \Sigma_{\sigma(C)} \to [0, 1]$, defined such that $p^*_{\mathsf{e}}(U) = \inf(\{\mathrm{fcap}_{\mathsf{e}}(A) \,|\, \sigma(A) \supseteq U\})$, is a measure. Therefore each ensemble of a probability context is associated with a unique measure over the spectrum.*

Conjecture 1.262. *A convex subset $U \subset \mathcal{E}$ is a classical probability context if and only if it is the affine subspace of probability measures of a vector space of measures over a sample space Ω.*

Conjecture 1.263. *Discrete and continuous classical ensemble spaces are classical probability contexts.*

A quantum ensemble space, however, does not allow a description in terms of classical probability precisely because it allows separate multidecompositions.

Proposition 1.264. *A quantum ensemble space is not a classical probability context.*

Proof. Let $\mathcal{E}$ be a quantum ensemble space. Let $A \subseteq \mathcal{E}$ the space of mixtures of a two dimensional subspace (i.e. a Bloch sphere) and consider the states x^+, x^-, z^+, and z^-, which are points on the surface that form a square. These are all separate ensembles. We have $\frac{1}{2}x^+ + \frac{1}{2}x^- = \frac{1}{2}z^+ + \frac{1}{2}z^-$ therefore the space allows separate multidecomposition and is not a classical probability context. □

However, given a maximal set of commuting observables, we can define a classical probability context as the set of all mixed states that commute with all the observables. This makes the spectrum of all the density operator coincide with the product of the spectra of all commuting observables, and therefore those mixed states can be characterized by a measure over the spectra.

Conjecture 1.265. *Let $\mathcal{E}$ be a quantum ensemble space. Let O_i a maximal set of commuting observables. Let $C \subseteq \mathcal{E}$ be the set of mixed state that commute with all O_i. Then C is a classical probability context.*

Alternative definition for contexts

Alternatively, a context can be define in terms of a lattice of orthogonal sets.

Definition 1.266. *Let $\mathfrak{L}$ be the lattice of subspaces and $\wedge, \vee, (\cdot)^\perp$ be respectively the join, meet and orthogonal complement. A **context** is a lattice of subspaces $C \subseteq \mathfrak{L}$ such that*

1. *if $\{X_i\}_{i\in I} \in C$ then $\bigwedge_{i\in I} X_i \in C$, $\bigvee_{i\in I} X_i \in C$ and $X_i^\perp \in C$*
2. *if $X, Y \in C$ and $X \cap Y = \varnothing$, then $X \perp Y$.*

That is, the join, meet and complement operation in $\mathfrak{L}$ and C are the same (i.e. smallest subspace that contains all, biggest subspace contained by all, orthogonal subspace). The additional property is that disjoint subspaces are orthogonal.

Proposition 1.267. *Let $C \subseteq \mathfrak{L}$ be a context. Then* scap *is an additive measure over the context. That is,* $\mathrm{scap}(X \vee Y) = \mathrm{scap}(X) + \mathrm{scap}(Y)$ *for all $X, Y, X \vee Y \in C$ such that $X \cap Y = \varnothing$.*

Proof. Since $X \vee Y$ contains both X and Y, we have $X \cup Y \subseteq (X \vee Y)$. By 1.293, $\mathrm{hull}(X \cup Y) \subseteq (X \vee Y)$. By □

Definition 1.268. *An ensemble $\mathsf{e} \in \mathcal{E}$ is **compatible** with a context $C \subseteq \mathfrak{L}$ if it is orthogonally decomposable into an X-maximal sequence and $X^\perp$-maximal sequence for any $X \in C$.*

Corollary 1.269. *Let $\mathsf{e} \in \mathcal{E}$ an ensembles compatible with a context $C \subset \mathfrak{L}$. Then $p_\mathsf{e}(X^\perp) = 1 - p_\mathsf{e}(X)$.*

Proof. Since e is compatible with C, we can write $\mathsf{e} = p_i \mathsf{x}_i + \lambda_i \mathsf{y} + \epsilon_i \mathsf{e}_i$ where $\mathsf{x}_i \in X$ and $\mathsf{y}_i \in X^\perp$ are maximal component sequences. Note that $p_\mathsf{e}(X^\perp) = 1 - p_\mathsf{e}(X)$ if and only if $\epsilon_i \to 0$. Suppose it didn't. Then every e_i would have a component that is neither in X or $X^\perp$. That would mean it is orthogonal from all the element of X and of $X^\perp$. But X contains all the elements disjuct from $X^\perp$ and vice-versa. Therefore such component cannot exist, $\epsilon_i \to 0$ and $p_\mathsf{e}(X^\perp) = 1 - p_\mathsf{e}(X)$. □

Conjecture 1.270. *Let $C \subseteq \mathfrak{L}$ be a context and $\mathsf{e} \in \mathcal{E}$ an ensemble compatible with the context. Then p_e is an additive measure over the context. That is, $p_\mathsf{e}(X \vee Y) = p_\mathsf{e}(X) + p_\mathsf{e}(Y)$ for all $X, Y, X \vee Y \in C$ and $X \cap Y = \varnothing$.*

Remark. The additivity of probability is more of a special condition than one may expect. Naively, one would expect that if U and V are separate, then $p_\mathsf{e}(U \cup V) = p_\mathsf{e}(U) + p_\mathsf{e}(V)$. That is, the component of e in $U \cup V$ is simply the sum of the components within U and V,

which are going to be separate and orthogonal (i.e. will not overlap). However this is not true in quantum mechanics.

Take a singe quibit and let e be the maximally mixed state. Then $p_{\mathsf{e}}(\{x\}) = \frac{1}{2}$ for any pure state x. However, if we take two pure states x and y that are not orthogonal, the maximally mixed state is not a mixture of x and y, which means $p_{\mathsf{e}}(\{x, y\}) < 1 = p_{\mathsf{e}}(\{x\}) + p_{\mathsf{e}}(\{y\})$.

Conjecture 1.271. *An ensemble space $\mathcal{E}$ is a classical ensemble space if and only if $\mathfrak{L}$ is a context.*

Conjecture 1.272. *An ensemble space $\mathcal{E}$ is orthogonally decomposable if and only if every ensemble is compatible with some context.*

1.16 Open problem: Topological Measures

Tags: Topology, Measure theory

We still need to understand what requirements a measure must satisfy to describe a well-posed physical problem. The issue is that the set of all probability measures defined over a space is too broad, but it is not clear how it should be restricted. There must be a link between the underlying experimental verifiability that motivates topologies and the possible outputs associated to the procedures that are connected to the measures.

If we consider the real line, given the Lebesgue measure μ, Lebesgue's decomposition theorem assures us that every measure ν can be divided into these three components: an absolutely continuous part ν_{ac} with respect to the Lebesgue measure (i.e. $\nu_{ac}(U) = 0$ for all sets U for which $\mu(U) = 0$), a pure point part ν_{pp} (i.e. there is a set of countable points $\{x_i\}$ such that $\nu_{pp}(\{x_i\}^{\mathsf{C}}) = 0$) and a singular continuous part ν_{sc} (i.e. $\mu(\{x\}) = 0$ for all x and there is a set U such that $\mu(U) = 0$ and $\nu_{sc}(U^{\mathsf{C}}) = 0$).

To make this more concrete, let us assume that we are working on phase space, μ is the Liouville measure, which in statistical mechanics quantifies the states in a region, and p is a probability measure. If p is absolutely continuous with respect to μ, it means that if a region has no states, it will have zero probability. This is the requirement under which the Radon-Nikodym theorem applies and a probability density $\frac{dp}{d\mu}$ exists. Without this requirement, for example, the entropy cannot even be defined. On physics grounds, this requirement makes a lot of sense. A pure point measure would correspond to the case where the probability is concentrated into a few isolated points. In physics, these are often represented with delta functions. There is an inherent unphysicality of these measures: if we really have a continuum, it would make no sense to say that we are are able to prepare an exact value with certainty. We are essentially saying that we are able to concentrate the whole distribution on a set that is not experimentally verifiable (i.e. it is a closed set with no interior) and contains no states (i.e. the Liouville measure is zero). Yet, there are some cases where the distribution is over real values, but only certain values are allowed. For example, the mass spectrum for particles or the energy spectrum for a bound quantum system can only take certain values. In this case, the Lebesgue measure is the one that is meaningless, because the region in between the allowed value contains no physically meaningful cases. A singular continuous measure may be something physically irrelevant, like the Cantor distribution which is defined only on a Cantor set, but it may also represent a constrained distribution. For example, a uniform distribution over the surface of a sphere, if defined over three dimensional space, would have support on a measure zero set, and have zero probability at every point. The issue, again, is that the

imposition of the constraint has to be applied to the whole mathematical structure, not just the probability measure.

While absolutely continuity is a requirement for a physically meaningful probability measure, it may not be enough. Consider a probability measure with a uniform distribution over a fat Cantor set. It would seem to be absolutely continuous but unlikely to be of physical interest.

Physically, the requirement of experimental verifiability poses constraints on what measures are possible, which means, mathematically, that there measures must satisfy some compatibility condition with the topology. In fact, it can be argued that some topological spaces may not allow any experimentally verifiable probability measure. Take, for example, the experimental domain generated by verifiable statements of the type *"there are at least i elements."* That is, we can only verify the existence of some objects (e.g. fundamental particles) and not able to verify the non-existence of others. This gives us the natural numbers with the right topology. Can we, in line of principle, put a measure on this space? The problem is that, experimentally, we can only exclude a range below what was verified, so we cannot perform a repeated set of measurement and build a histogram. The only thing we would be able to confirm is the case that below a certain threshold we have probability zero.

It is not clear, however, what is the general rule. Naively, we thought that, given a verifiable statement, the non-termination condition (i.e. the boundary of the corresponding open set) cannot be assigned a non-zero measure. This intuition was reinforced by the finding that open sets have the feature that their boundary is nowhere dense (i.e. it has no interior), and sets with a nowhere dense boundary actually form a nice algebra. However, as we see in 1.223, there exists open sets that have boundaries with non-zero measures. Therefore, finding the right characterization is still an open problem.

We leave here the characterization of the algebra of sets with a nowhere dense boundary, in case it may be useful.

Nicely boundaried sets

Part of an original attempt to square measure and topology, sets with a nowhere dense boundary form a nice algebra that may, or may not, be useful.

Definition 1.273. *Let X be a topological space. A set $A \subseteq X$ is* ***nicely boundaried*** *if its boundary is nowhere dense.*

Proposition 1.274. *The following are true:*

1. *open sets are nicely bounded*
2. *the complement of a nicely bounded set is nicely bounded*
3. *closed sets are nicely bounded*
4. *the union of nicely bounded sets is nicely bounded*
5. *the intersection of nicely bounded sets is nicely bounded.*

Open and closed sets are nicely bounded. The complement of a nicely bounded set is nicely bounded. The intersection and the union of nicely bounded sets are nicely bounded.

The closure of open sets under complement, finite intersection and finite union is a Boolean algebra R_X of nicely bounded sets.

Proof. For 1, let $A \subseteq X$ be an open set. We have $A = \text{int}(A)$ and, since ∂A and $\text{int}(A)$ are disjoint, ∂A and A are disjoint. Let $U \subseteq \partial A$ be an open set. Then U is disjoint from A and therefore $U \subseteq \text{ext}(A)$. Since the boundary and the exterior are disjoint, U must be the empty set.

For 2, let A be nicely bounded. Since $\partial A = \partial A^{\mathsf{C}}$, A^{C} is also nicely bounded.

For 3, note that closed sets are complements of open sets, which are nicely bounded.

For 4, let $A, B \subseteq X$ be two nicely bounded sets. The boundary satisfies the property $\partial(A \cup B) \subseteq \partial A \cup \partial B$. The union of two nowhere dense sets is nowhere dense and a subset of a nowhere dense set is nowhere dense. Therefore $\partial(A \cup B)$ is nowhere dense and $A \cup B$ is a nicely bounded set.

For 5, since complements and unions of nicely bounded sets are nicely bounded, by De Morgan intersections are nicely bounded as well. □

Proposition 1.275. *Let X be a topological space, the closure of open sets under complement, finite intersection and finite union is a Boolean algebra R_X of nicely bounded sets.*

Proof. By definition, R_X is closed under complement, finite intersection and finite union and is therefore a Boolean algebra. Since open sets are nicely bounded, and the operation preserve the property, R_X is composed of only nicely bounded sets. □

Remark. Note that the algebra R_X does not contain all the nicely bounded sets. Consider a circle in $\mathbb{R}^2$. Let U be the open set of all the interior points and B be the points on the circle at a rational angle. Now consider $A = U \cup B$: U is the interior of A and B its boundary. The boundary is nowhere dense, therefore A is a nicely bounded set. However, we cannot generate A with finite operations from open sets of $\mathbb{R}^2$.

1.17 Open problem: Spectrum of a quantity

Tags: Ensemble spaces, Spectral theory, Topological vector spaces, Order theory

We are looking for a definition of the spectrum of a quantity (i.e. linear functional) that can be defined for a general ensemble space (i.e. convex space). It would already be helpful to establish whether the entropic structure is required, or whether the convex structure is sufficient.

We leave here the latest attempt, more for inspiration and reference.

Definition 1.276. *Let $\mathcal{E}$ be an ensemble space and let $F : \mathcal{E} \to \mathbb{R}$ be a statistical quantity. Let $x, y \in \mathbb{R}$ such that $x \leq y$. We say a is **within** $[x, y]$ **of** F if $F(\mathsf{e}) \in [x, y]$ for every component e of a. Analogous definitions apply for intervals bounded only on one side.*

*We say $\mathsf{a} \in \mathcal{E}$ is **supported by** $[x, y]$ **of** F if:*

1. *a is within $[x, y]$ of F*
2. *let e be in the flat generated by a and all ensembles within $(-\infty, x]$, then $F(\mathsf{e}) \leq F(\mathsf{a})$*
3. *let e be in the flat generated by a and all ensembles within $[y, +\infty)$, then $F(\mathsf{a}) \leq F(\mathsf{e})$.*

Analogous definition applies for intervals bounded only on one side.

*An **eigenstate** of F is an ensemble $\mathsf{a} \in \mathcal{E}$ that is supported by a singleton $U = [a, a]$, where $a \in \mathbb{R}$ is called the respective **eigenvalue**.*

Proposition 1.277. *Let $\mathcal{E}$ be a discrete classical ensemble space over the set of points X, and F a statistical quantity. Then $F(X)$ is the set of possible eigenvalues, and an eigenstate is any ensemble that can be expressed as a mixture of points with the same eigenvalue.*

Proof. First we show that any extreme point $\mathsf{a} \in \mathcal{E}$ is an eigenstate with eigenvalue $F(\mathsf{a})$. First, we have $F(\mathsf{a}) \in [F(\mathsf{a}), F(\mathsf{a})]$, which satisfies the first condition for the support. Now let $\mathsf{b} \in \mathcal{E}$ be an ensemble within $(-\infty, F(\mathsf{a})]$. Let $\{\mathsf{e}_i\}$ be the set of extreme points that are components of b. Since $\{\mathsf{e}_i\}$ are components of b, $F(\mathsf{e}_i) \leq F(\mathsf{a})$ since b is within $(-\infty, F(\mathsf{a})]$. Recall that orthogonality, for a classical space, coincides with disjoint support of the probability distribution. Also recall that, in a discrete classical ensemble space, all ensembles can be decomposed in terms of the extreme points. Then $\langle\{\mathsf{a}, \mathsf{b}\}\rangle_\perp$ is exactly $\text{hull}(\{\mathsf{e}_i\} \cup \{\mathsf{a}\})$. Given the linearity of F, for any convex combination e of e_i and a we have $F(\mathsf{e}) \leq F(\mathsf{a})$. This satisfies the second condition for the support. The argument can be repeated with the upper bound, so the third condition is also satisfies. This means that any extreme point is a is an eigenstate of F with eigenvalue $F(\mathsf{a})$.

Note that the above argument works also if e is a mixture of eigenstates with the same eigenvalues.

Lastly, we show that any ensemble that is a mixture of two extreme points with different eigenvalue is not an eigenstate. Let $\mathsf{e} = p\mathsf{a} + \bar{p}\mathsf{b}$ be a mixture of two extreme points $\mathsf{a}, \mathsf{b} \in \mathcal{E}$ such that $F(\mathsf{a}) \neq F(\mathsf{b})$. Without loss of generality, suppose $F(\mathsf{a}) < F(\mathsf{b})$. Given the linearity of F, we have $F(\mathsf{a}) < F(\mathsf{e}) < F(\mathsf{b})$. Therefore e is not within $[F(\mathsf{e}), F(\mathsf{e})]$, is not supported by $[F(\mathsf{e}), F(\mathsf{e})]$ and therefore is not an eigenstate. This also means that any ensemble of extreme points with different eigenvalue is not an eigenstate. □

Proposition 1.278. *Let $\mathcal{E}$ be a continuous classical ensemble space over the symplectic manifold X, and F a statistical quantity. Then $\rho \in \mathcal{E}$ is supported by $[x, y]$ of F if and only if the $\int_{[x,y]} \rho d\mu = 1$. Moreover, $a \in \mathbb{R}$ is an eigenvalue if and only if there is an open set $U \subset X$ such that $f^{-1}(a) \supseteq U$ where $f : X \to \mathbb{R}$ is the state variable corresponding to F.*

Proof. Let F be a statistical quantity over a continuous classical ensemble space. Then we can find $f : X \to \mathbb{R}$ such that $F(\rho) = \int_X f\rho d\mu$ for all $\rho \in \mathcal{E}$. Given an interval $[x, y] \subseteq \mathbb{R}$, then we can define $U_{[x,y]} = f^{-1}([x, y])$ which can be understood as the set of particle states for which the value of the property is within the bounds. Any probability distribution ρ can be decomposed as

$$\rho = p_{(-\infty,x]}\rho_{(-\infty,x]} + p_{[x,y]}\rho_{[x,y]} + p_{[y,-\infty)}\rho_{[y,-\infty)}$$

where

$$\begin{aligned} p_{(-\infty,x]} &= \int_{U_{(-\infty,x]}} \rho d\mu \quad p_{[x,y]} = \int_{U_{[x,y]}} \rho d\mu \quad p_{[y,-\infty)} = \int_{U_{[y,-\infty)}} \rho d\mu \\ \rho_{(-\infty,x]} &= \frac{\rho|_{U_{(-\infty,x]}}}{p_{(-\infty,x]}} \qquad \rho_{[x,y]} = \frac{\rho|_{U_{[x,y]}}}{p_{[x,y]}} \qquad \rho_{[y,-\infty)} = \frac{\rho|_{U_{[y,-\infty)}}}{p_{[y,-\infty)}} \end{aligned} \tag{1.279}$$

Suppose the support of ρ is within $U_{[x,y]}$. Then $\int_X \rho d\mu \in [x, y]$. Since any of its components will also have support within $U_{[x,y]}$, the expectation of f for all its components is also within

$[x, y]$, within means that ρ is within $[x, y]$ of F. Conversely, if the support of ρ is not within $U_{[x,y]}$, then we can find a component whose support is either within $U_{(-\infty,x]}$ or $U_{[y,-\infty)}$. Therefore, ρ is within $[x, y]$ of F if and only if the support of ρ is within $U_{[x,y]}$.

Now suppose ρ is within

Now we show that ρ is supported by $[x, y]$ of F if $\rho_f(f) = 0$ for all $f \notin [x, y]$ where ρ_f is the probability density function (PDF) of F. Since ρ is a probability measure and f is a random variable, we can define a PDF ρ_f. This can also be understood as the marginal of ρ over f. If ϕ is a component of ρ, the corresponding ϕ_f will be a component of ρ_f. If ρ is supported by $[x, y]$ of F, then the expectation of f for any component of ρ_f must be within $[x, y]$. This can only happen if the support of ρ_f is within $[x, y]$, as any component below or above those bounds would have expectation below or above the bounds. Conversely, if the support of ρ_f is within $[x, y]$ Therefore, ρ is within $[x, y]$ of F if $\rho_f(f) = 0$ for all $f \notin [x, y]$ where ρ_f □

1.18 Open problem: Conditional expectation values

Tags: Ensemble spaces, Probability, Non-additive measures

There should be a way to generalize the notion of conditional expectation to a generic ensemble space. The general idea should be that, given a target ensemble and a set of ensembles, we find the biggest component of the target that is a mixture of ensembles of the set. The expectation of a quantity for the target ensemble restricted to that set would be the quantity evaluated on the representative. In classical probability, it would recover the expectation of a random variable conditioned to an event.

Maximal component sequence

Definition 1.280. *Let* $\mathsf{e} \in \mathcal{E}$ *be an ensemble. A sequence of ensembles* $\{\mathsf{a}_i\} \subseteq \mathcal{E}$ *is an* ***increasing component sequence of*** e *if we can write*

$$\mathsf{e} = p_i \mathsf{a}_i + \bar{p}_i \mathsf{b}_i$$
$$\mathsf{a}_{i+1} = \frac{p_i}{p_{i+1}} \mathsf{a}_i + \frac{p_{i+1} - p_i}{p_{i+1}} \Delta \mathsf{a}_i$$

where $\{\mathsf{b}_i\} \subseteq \mathcal{E}$, $\{\Delta \mathsf{a}_i\} \subseteq \mathcal{E}$ *and* $\{p_i\} \subseteq (0, 1]$ *is an increasing sequence. The* ***fraction*** *of the sequence is the limit* $p_i \to p$.

Remark. Since the sequence of p_i is increasing and is bounded, it must converge. The set of all possible limits is bounded and therefore must have a supremum.

Definition 1.281. *Let* $\mathsf{e} \in \mathcal{E}$ *be an ensemble and* $A \subseteq \mathcal{E}$ *a set of ensembles. Then the* ***A-components of*** e *are the components of* e *that are mixtures of* A. *That is,* $A_{\mathsf{e}} = \{\mathsf{e}_1 \in \mathrm{hull}(A) \mid \exists p \in (0, 1], \mathsf{e}_2 \in \mathcal{E} \text{ s.t. } \mathsf{e} = p\mathsf{e}_1 + \bar{p}\mathsf{e}_2\}$. *An* ***increasing A-component sequence of*** e *is an increasing component sequence of* e *such that* $\{\mathsf{a}_i\} \subseteq \mathrm{hull}(A)$ *and* $\{\Delta \mathsf{a}_i\} \subseteq \mathrm{hull}(A)$. *The sequence* $\{\mathsf{a}_i\} \subseteq \mathrm{hull}(A)$ *is* ***maximal*** *if the fraction of the sequence is* $\mathrm{fcap}_{\mathsf{e}}(A)$.

Corollary 1.282. *The fraction of A-component sequences are bounded by* $\mathrm{fcap}_{\mathsf{e}}(A)$.

Proof. For all elements of component sequences we have $p_i \leq \mathrm{fcap}_{\mathsf{e}}(A)$. Therefore the limit of p_i cannot exceed $\mathrm{fcap}_{\mathsf{e}}(A)$. □

Proposition 1.283. *Let* $\mathsf{e} \in \mathcal{E}$ *be an ensemble and* $A \subseteq \mathcal{E}$ *a set of ensembles, then there exists a maximal* A*-component sequence of* e*.*

Proof. Consider the hull of A. This can be ordered by the fraction frac_{e}, which is less or equal to one. If there is a maximum, then simply take a sequence of an element with the maximum faction. If not, we can take a sequence of ever increasing fractions whose limit is the fraction capacity, and find a corresponding sequence of ensembles within $\text{hull}(A)$. By definition, this will be a maximal A-component sequence of e. □

Proposition 1.284. *Let* $\{\mathsf{a}_i\} \subseteq \text{hull}(A)$ *be an* A*-component sequence of* $\mathsf{e} \in \mathcal{E}$*. A* ***complement*** *sequence is an increasing sequence* $\{\mathsf{b}_i\} \subseteq \mathcal{E}$ *such that*

$$\mathsf{e} = p_i \mathsf{a}_i + q_i \mathsf{b}_i + \overline{(p_i + q_i)} \epsilon_i$$

where $\{\epsilon_i\} \in \mathcal{E}$ *and* $(p_i + q_i) \to 1$*. In this sense, the sum of the two sequences converges to* e*.*

Remark. Note that complement sequences always exist since we can set $\epsilon_i = \mathsf{b}_i$, $q_i = \bar{p}$ and recover the definition of a component sequence.

Proposition 1.285. *An* A*-component sequence* $\{\mathsf{a}_i\} \subseteq \text{hull}(A)$ *of* $\mathsf{e} \in \mathcal{E}$ *is maximal if and only if it admits a complement sequence that is always separate from the hull of* A*. That is,* $\{\mathsf{b}_i\} \pitchfork \text{hull}(A)$*.*

Proof. Suppose that $\{\mathsf{a}_i\}$ is maximal and let $\{\mathsf{b}_i\}$ be a complement.

$$\mathsf{e} = p_i \mathsf{a}_i + q_i \mathsf{b}_i + \overline{(p_i + q_i)} \epsilon_i.$$

□

Proposition 1.286. *Let* $\mathsf{e} \in \mathcal{E}$ *be an ensemble,* $A \subseteq \mathcal{E}$ *a set of ensembles and* a_i *and* b_i *two maximal* A*-component sequence of* e*. Then we can write* $\mathsf{a}_i = p_i \mathsf{b}_i + \bar{p}_i \mathsf{c}_i$ *where* $\mathsf{c}_i \in \mathcal{E}$*,* $p_i \in [0, 1]$ *and* $p_i \to 1$*.*

Proof. Note that we can always write $\mathsf{a}_i = p_i \mathsf{b}_i + \bar{p}_i \mathsf{c}_i$ because we can always choose $p_i = 0$ and $\mathsf{c}_i = \mathsf{a}_i$. Therefore, if the proposition is not true, we can always find $p_i \to p$, except that $p \neq 1$.

Suppose the proposition is not true. We can still write $\mathsf{e} = \lambda_i \mathsf{a}_i + \bar{\lambda}_i \mathsf{d}_i = \lambda_i p_i \mathsf{b}_i + \lambda_i \bar{p}_i \mathsf{c}_i + \bar{\lambda}_i \mathsf{d}_i$. But b_i is maximal, therefore $\lambda_i p_i \mathsf{b}_i$ can be increased. But $\mathsf{c}_i \in \text{hull}(A)$, which means we can write an A-component sequence whose fraction is higher than the maximal, which cannot be. Therefore the proposition is true.

TODO: this may need to be fixed. Can't reconstruct the argument. □

Conjecture 1.287. *Let* $e \in \mathcal{E}$ *be an ensemble,* $F : \mathcal{E} \to \mathbb{R}$ *be a statistical variable and* $A \subset \mathcal{E}$ *be a set of ensemble. Let* a_i *and* b_i *be maximal* A*-component sequences of* e*. Then* $\lim_{i \to \infty} F(\mathsf{a}_i) = \lim_{i \to \infty} F(\mathsf{b}_i)$*.*

Definition 1.288. *Let* $e \in \mathcal{E}$ *be an ensemble,* $F : \mathcal{E} \to \mathbb{R}$ *be a statistical variable and* $A \subset \mathcal{E}$ *be a set of ensemble. Then the* ***contribution to*** F ***over*** A ***given*** e *is the limit of variable for a maximal* A*-component sequence of* e*. That is,* $F_{\mathsf{e}}(A) = \lim_{i \to \infty} F(\mathsf{a}_i)$ *where* a_i *is a maximal* A*-component sequence of* e*.*

Proposition 1.289. *The contribution to F is a set function.*

Proof. Since F_{e} takes a set of an argument and returns a real value, is a set function. □

Remark. Note that the contribution cannot be monotone in the same way that the expectation of a random variable to an event is not monotone. Not clear whether the additivity of the variable may tell us something about the additivity of the contribution.

Conjecture 1.290. *We can define a "derivative" f_{e} between F_{e} and $\mathrm{fcap}_{\mathsf{e}}$ such that $F_{\mathsf{e}}(A) = \int_A f_{\mathsf{e}}\, d\mathrm{fcap}_{\mathsf{e}}$ where $\int$ is the ??? integral.*

Remark. This may still not be the correct formulation of the problem. Note that A is a set of ensembles. In the classical case it would correspond to a set of probability measures, not a subset of the sample space (i.e. an event). However, in the discrete case and in the quantum case, A can also be restricted to a set of pure state (i.e. extreme points), which would correspond to a subset of the sample space. It may be worth at least understanding that case.

1.19 Open problem: Ensemble subspaces

Tags: Ensemble spaces, convex spaces

We should create a notion of ensemble subspaces that recovers those of classical and quantum mechanics. In classical probability, each event identifies the subspace of probability measures whose support is within that even. In quantum mechanics, each subspace of the Hilbert space identify a subspace of density operators that can be defined on that subspace alone.

Given that orthogonality of subspaces in all three cases (i.e. discrete/continuous classical and quantum) corresponds to ensembles saturate the upper entropy bound (i.e. orthogonal in the sense of the entropy), we will define the notion of subspaces based on the entropy.

We will now use the previous results where X is an ensemble space and R is the orthogonality relation between two ensembles.

Proposition 1.291. *Let $\mathcal{E}$ be an ensemble space, orthogonality $\perp$ is an irreflexive symmetric relation.*

Proof. Since any ensemble mixed with itself saturates the lower bound of entropy, it does not satureate the upper bound and it is therefore not orthogonal with itself. Therefore orthogonality is irreflexive. Two elements are orthogonal if they satureate the upper entropy bound. This does not depend on their order. Therefore orthogonality is symmetric. □

Definition 1.292. *Let $\mathcal{E}$ be an ensemble space and $X \subseteq \mathcal{E}$ be a subset. The **orthogonal complement** $X^\perp \subseteq \mathcal{E}$ is the set of all ensembles that are orthogonal from all elements of X. An **ensemble subspace** is a subset $X \in \mathcal{E}$ such that $X = (X^\perp)^\perp$.*

Proposition 1.293. *Let $X \subseteq \mathcal{E}$ be an ensemble subspace. Then $\mathrm{hull}(U) \subseteq X$ for all $U \subseteq X$.*

Proof. Let $U \subseteq X$. Then $U \perp X^\perp$. But since mixtures preserve orthogonality, we also have $\mathrm{hull}(U) \perp X^\perp$. Therefore $\mathrm{hull}(U) \subseteq X$. □

Corollary 1.294. *Let $X \subseteq \mathcal{E}$ be an ensemble subspace, then $\mathrm{hull}(X) = X$.*

Proof. Since $X \subseteq X$, $\text{hull}(X) \subseteq X$ by 1.293. By 1.38, we also have $X \subseteq \text{hull}(X)$. Therefore $\text{hull}(X) = X$. □

Remark. The converse is not true: $\text{hull}(X) = X$ does not mean X is an ensemble subspace. For example, let $\mathcal{E}$ the ensemble space of a two state quantum system (i.e. the Bloch ball). Let U be the set of all mixtures of two pure states (i.e. and the segment that connects two points on the surface). Then U is closed under mixture (i.e. convex combinations) but it is not a subspace. In fact we have that $U^\perp = \varnothing$ and therefore $(U^\perp)^\perp = \mathcal{E}$.

Proposition 1.295. *Let $\mathcal{E}$ be a discrete classical theory and $X \subseteq \mathcal{E}$ an ensemble subspace. Then X is the set of all probability distributions over a subset A of pure states.*

Proof. Let $\mathcal{E}$ be a discrete classical ensemble space. Then each ensemble is a sequence p_i such that $\sum_i p_i = 1$. The dimensionality of the space fixes the range of i. Note that, given that the space is discrete, the collection p_i can be understood as a continuous function of the pure elements.

Note that two probability distributions will be orthogonal if and only if their supports are disjoint. Therefore orthogonality for a discrete classical ensemble space is the same as having disjoint support. This means that a subspace is given by all possible probability distributions over a subset A of all pure states. □

Proposition 1.296. *Let $\mathcal{E}$ be a continuous classical theory and $X \subseteq \mathcal{E}$ an ensemble subspace. Then X is a set of measures whose support is a regular closed set (i.e a set that is the closure of its own interior).*

Proof. Let $\mathcal{E}$ be a continuous classical ensemble space. Then the probability density associated to each ensemble is a continuous function over a symplectic manifold M that integrates to one. Since the function is continuous, it is non-zero on an open set, and the support is the closure of that set.

Consider two continuous probability densities. They will be orthogonal if and only if they have disjoint support, meaning the interior of their supports are disjoint. Therefore orthogonality for a continuous classical ensemble space is the same as having disjoint support.

Take a set of ensembles $U \subseteq \mathcal{E}$. An ensemble $\mathsf{e} \in \mathcal{E}$ will be orthogonal from all ensembles in U if and only if its support is disjoint from the union of all the supports of all the elements of U. Therefore $U^\perp$ is the set of all ensembles whose support is within the closure of of the exterior of the union of all supports of elements of U. With a similar logic, $(U^\perp)^\perp$ is the set of all ensembles whose support is within the closure of the exterior of the union of the support of elements of $U^\perp$. Therefore, X will contain all probability measure whose support within a set $A \subseteq M$ that is the that $A = \overline{\text{int}(\text{int}(A))}$ and is therefore a regular set.

Now take a regular closed set $A \subseteq M$ and let X be the set of all continuous probability densities whose support is within A. Then $X^\perp$ is the subspace of all continuous probability densities that have support within $\overline{\text{ext}(A)}$ and $(X^\perp)^\perp$ is the subspace of all continuous probability densities that have support within $\overline{\text{ext}(\text{ext}(A))} = \overline{\text{int}(A)} = A$. Therefore X is a subspace. □

Proposition 1.297. *Let $\mathcal{E}$ be a quantum ensemble space and $X \subseteq \mathcal{E}$ an ensemble subspace. Then X is the set of mixed state whose support is a subspace of the corresponding Hilbert space.*

Proof. Let $\mathcal{E}$ be a quantum ensemble space. Then each ensemble is a density operator defined over a Hilbert space $\mathcal{H}$. Consider two density operators. They will be orthogonal, that is the entropy increase is maximal, if and only if they are defined on orthogonal subspaces of $\mathcal{H}$. That is, $\mathsf{e}_1|\psi\rangle \neq 0$ only if $\mathsf{e}_2|\psi\rangle = 0$ and vice-versa. Therefore X contains exactly all the mixed states whose support is a subspace of $\mathcal{H}$. □

1.20 Open problem: Ensemble space composition

Tags: Ensemble spaces, convex spaces

We need a definition for composite systems that takes two ensemble spaces and create the product. The issue is that it is unclear whether this can be a single definition, as product spaces for classical ensembles and quantum ensembles are different. Note, however, that the additional structure given by the entropy and, potentially, by the missing Lie algebraic structure may fill the gaps.

1.21 Open problem: Poisson structure over ensemble spaces

Tags: Ensemble spaces, Poisson structure, Lie algebras

Both classical and quantum mechanics contain a Poisson/symplectic structure implemented by Poisson brackets and commutators. We need to be able to generalize this structure on ensemble spaces, without reference to the classical and quantum implementation. The goal would be to write generalized Hamiltonian equations that work in both cases.

Part III

Blueprints for the work ahead

This part is dedicated to preliminary ideas and work in progress, meaning that it includes conjectures, current thinking and approaches that may or may not turn out to be good ideas. As the work matures, sections and chapters from this part will be merged into the main part.

The first chapter will give an overview of rough general ideas within reverse physics. The second chapter will do the same for physical mathematics. The subsequent chapters capture an in-depth stab at a particular problem that, if solved, it would likely turn out to be a standalone chapter in either reverse physics or physical mathematics.

Chapter 1

Reverse Physics

1.1 Classical mechanics

The work on classical mechanics is considered mostly concluded, in the sense that suitable initial assumptions have been identified. There are still a few open issues, such as the case of variable mass, the generalization of the directional DOF to the relativistic case, or clarifying the nature or the generalization to infinite DOFs (i.e. field theories).

Curvature for particle dynamics

The assumption of kinematic equivalence already gives us relativistic Hamiltonians. Does it also give us a relationship between the curvature of the metric tensor and the forces acting on the particles?

The setup is the following. Suppose we have two vectors in the extended phase space $d\xi^a = \{dq^\alpha, 0\}$ and $d\nu^a = \{0, dp_\alpha\}$. Using the symplectic form we have the invariant $d\xi^a \omega_{ab} d\nu^b = dq^\alpha dp_\alpha$. Under the kinematic assumption we have $dq^\alpha = dx^\alpha$ and $dp_\alpha = m g_{\alpha\beta} du^\beta + \mathfrak{q} A_\alpha$. We have $d\xi^a \omega_{ab} d\nu^b = dx^\alpha m g_{\alpha\beta} du^\beta + dx^\alpha \mathfrak{q} A_\alpha$.

Since the two terms have to match at each point and the symplectic form has the same components at each point, can we constrain the change of the components of $g_{\alpha\beta}$? The general idea would be that components of $g_{\alpha\beta}$ may have to change in space/time coordinately with A_α as to make $d\xi^a \omega_{ab} d\nu^b = dq^\alpha dp_\alpha$ remain the same. Note that derivatives in q^α are taken at constant p_α while derivatives in x^α are taken at constant u^α.

1.2 Thermodynamics

Process entropy

The key to recover thermodynamics is finding a definition of entropy that applies in very general cases and recovers the usual definition. Instead of using the logarithm of the count of states, we use the logarithm of the count of possible evolutions. That is, the ways a system can evolve under a specific process. The entropy of the system is automatically relativistic (i.e. we are essentially counting "worldlines" of the overall system in its state space) and is process dependent (i.e. contextual).

In the case of deterministic and reversible evolution, the count of states is equivalent to the count of evolutions, and therefore the usual definition is recovered. In the case of stochastic steady state over continuous time, that is when the probability distribution stabilizes, the

states will traverse infinitely many states within a small time difference dt. The count of evolutions, then, can be shown to reduce to the permutations of infinite sequences which recovers the Gibbs/Shannon entropy.

As for the behavior of entropy, the idea is that for a specific process, the state at a particular time identifies a set of possible evolutions. This would be the entropy of that state. Over the continuum, where states are points, the entropy would become a density of the count of evolutions. In essence, the entropy of a system at a particular time tells us how much or how little the evolution is constrained. In other words, it tells us how much the system is expected to fluctuate. As time evolves, the state changes, and the count of evolutions changes as well. If the evolution is deterministic, the evolutions can never split, in the sense that all the evolutions that end up in a particular state must all go to another state. This means that for a deterministic process the count can never decrease. If the evolution is deterministic and reversible, then the count must stay the same. This recovers the feature of entropy to be a non-decreasing quantity, which is conserved during reversible processes.

If the evolution allows equilibria, the evolutions will concentrate around states of equilibria. Given that states cannot go out of equilibrium once it has been reached, the count of evolutions is maximized at equilibrium. This recovers another feature of entropy.

Lastly, if two systems are independent, the way one evolves does not constrain the other. The total count of evolutions, then, is simply the product of the count of evolutions of the two systems. Since the entropy is the logarithm of the count of evolutions, it sums over independent systems. This recovers the last property of entropy.

Equation of state

If we study the space of equilibria, each state will have a well defined entropy. Therefore we have an equation of state $S(\xi^a)$ where ξ^a form a set of variables that fully identify the state. Moreover, as noted before, entropy is additive under system composition of independent systems.

In a process with equilibria different evolutions must converge to the same final state, which means the process entropy increases and is maximized at equilibria. This gives the general idea that entropy increases during an irreversible process. These results are therefore valid in general, no matter what type of system is being described.

To find thermodynamics specifically, we need an additional set of assumptions. First, all states are equilibria. Second, all state variables ξ^a are additive under system composition. Third, one of them, which we call internal energy U, is conserved under any evolution, including irreversible evolution. We can then write the equation of state as $S(U, x^i)$ and define the following quantities:

$$\begin{aligned} \frac{\partial S}{\partial U} &= \beta = \frac{1}{k_B T} \\ \frac{\partial S}{\partial x^i} &= -\beta X_i \end{aligned} \tag{1.1}$$

We can then express the differentials as:

$$\begin{aligned} dS &= \frac{\partial S}{\partial U} dU + \frac{\partial S}{\partial x^i} dx^i \quad = \beta dU - \beta X_i dx^i \\ dU &= T k_B dS + X_i dx^i \end{aligned} \tag{1.2}$$

This is essentially Gibbs' approach to thermodynamics.

Thermodynamic laws

To recover the laws, we need a few more definitions. We define a reservoir R as a system for which the internal energy U_R is the only state variable and the state entropy S_R is a linear function of U_R. That is, $\frac{\partial S_R}{\partial U_R} = \beta_R = \frac{1}{k_B T_R}$ is a constant. We call heat $Q = -\Delta U_R$ the energy lost by the reservoir during a transition.

We define a purely mechanical system M as a system for which the state entropy is zero for each state. That is, $S_M(U_M, x_M^i) = 0$. We call work $W = \Delta U_M$ the energy acquired by a purely mechanical system during a transition.

Now, consider a composite system made of a generic system A, a reservoir R and a purely mechanical system M. Consider a transition where we go to a new equilibrium. Since energy is additive under system composition, let us call U the total energy. Since energy is conserved we have:

$$\begin{aligned} \Delta U = 0 &= \Delta U_A + \Delta U_R + \Delta U_M = \Delta U_A - Q + W \\ \Delta U_A &= Q - W \end{aligned} \tag{1.3}$$

Since entropy is extensive, let us call S the total entropy. Since the process is going to an equilibrium, the entropy can only increase. We have:

$$\begin{aligned} 0 \leq \Delta S &= \Delta S_A + \Delta S_R + \Delta S_M = \Delta S_A + \beta_R \Delta U_R + 0 = \Delta S_A + \frac{-Q}{k_B T_R} \\ k_B \Delta S_A &\geq \frac{Q}{T_R} \end{aligned} \tag{1.4}$$

1.3 Quantum mechanics and irreducibility

Quantum mechanics can be recovered by swapping reducibility with irreducibility as shown in diagram 1.1, which can be used as a guide throughout this section.

The assumptions lie on the left column. Each assumption leads to one or two key insights that progressively lead to the physical concepts in the middle column. Each of these is then mapped to its corresponding formal framework on the right. Note that "quasi-static process" and "conserved density" both independently lead to the same result of "unitary evolution".

Irreducibility

The state space of quantum mechanics can be recovered under the:

Assumption V (Irreducibility)**.** *The state of the system is irreducible. That is, giving the state of the whole system says nothing about the state of its parts.*

Under this assumption the state of the system is automatically an ensemble over the state of the parts as preparation of the whole leaves the parts unspecified. For the same reason, the entropy of these ensembles must be the same, or some ensembles would provide more or less information about the parts. The whole task, then, is to characterize these ensembles without making specific assumptions on the parts.

Let $\mathcal{C}$ be the state space of the irreducible system. Let us call fragment a part of the irreducible system. The state of a fragment will be associated with a random variable uniformly distributed over the possible fragment states. As discussed in the context of classical

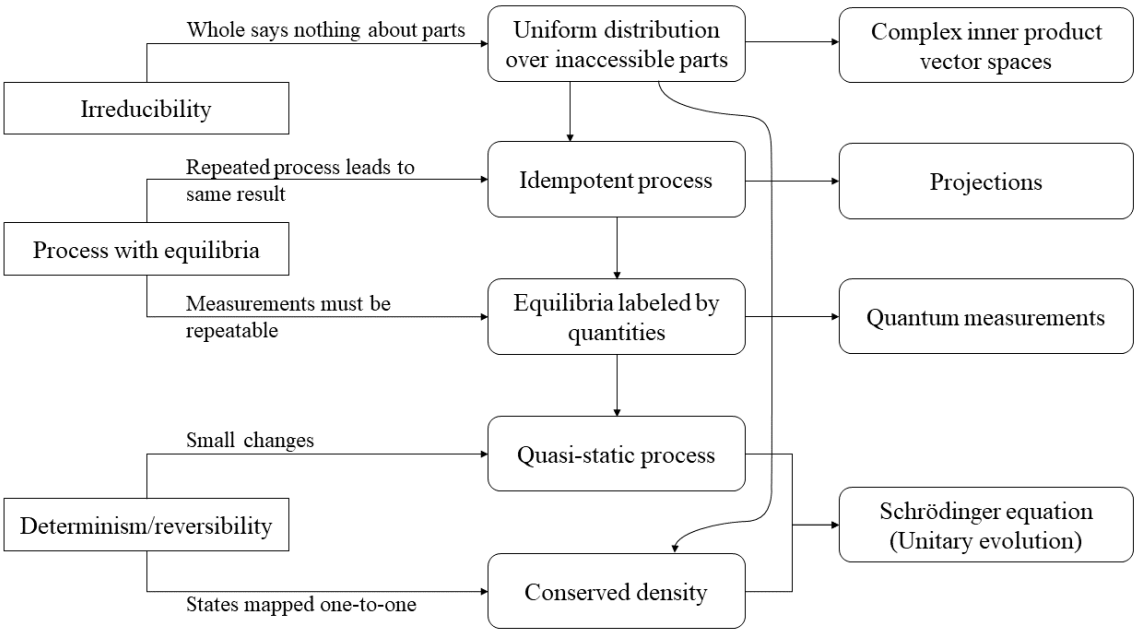


Figure 1.1: Assumptions for quantum mechanics

mechanics, distributions over states must be invariant and symplectic manifolds are the only manifolds over which invariant distributions can be defined. As we cannot say anything about the state of the fragments, the dimensionality of this manifold must be irrelevant as long as it is even dimensional. For simplicity, we can choose a two-dimensional one. Therefore we are interested in the space of bi-dimensional uniform distributions formed by a pair of two random variables A and B.

The values of the variables themselves are not relevant, as they are not physically accessible by assumption. However, the size of the system $\mu = \int \rho dA \wedge dB$ is relevant. Without loss of generality, we can rescale A and B such that the density ρ is not only uniform but unitary: $\rho = 1$. This way the size of the system is directly proportional to the area covered by the random variables. In other words: the more fragments there are, the more each fragment can swap its state with another without changing the whole, the more uncertainty there is on the state of the fragment, the higher the variance of the random variables.

Since only linear transformations will preserve the uniform distribution, we look to those. These are translations, stretches and rotations. Translations do not lead to other physically distinguishable states since the exact values of A and B are not physically accessible. Stretching of the distribution will correspond to an increase of the size of the system, which is physically accessible. However, only the stretching of the area is of interest. So, without loss of generality, we can set $\sigma_A = \sigma_B = \sigma$ and we have $\mu \propto \sigma^2$. Rotations just change the correlations which, by themselves, are not physically accessible. However, under addition the correlations still result in differences in variance and, indirectly, the size of the system, and therefore are physically interesting. The space of transformations is therefore given by two

parameters a and b such that:

$$\begin{aligned} C &= aA + bB \\ D &= -bA + aB \end{aligned} \tag{1.5}$$

Equivalently, we can use the complex number $c = a + \imath b$ to characterize the transformation, which we can note as $\tau(c)$. The increase/decrease in size is given by $a^2+b^2 = (a-\imath b)(a+\imath b) = c^*c$ and the change in correlation is given by the Pearson correlation coefficient $\rho_{A,\tau(c)A} = \cos\arg c$.

Putting it all together, we can characterize the state space $\mathcal{C}$ with a complex vector space. The linear combination represents the mixing of the different stochastic descriptions. Two vectors that only differ by a total phase are physically equivalent since a global change of correlation does not change the distribution.

We can define a scalar product $\langle\cdot|\cdot\rangle$ where the square norm induced corresponds to the size of the system (or equivalently to the strength of the random variable) and the phase difference corresponds to the correlations (the Pearson correlation coefficient). To see this, note the formal equivalence between the variance and norm rules under linear composition:

$$\begin{aligned} \sigma^2_{X+Y} &= \sigma^2_X + \sigma^2_Y + 2\,\sigma_X\sigma_Y\rho_{X,Y} \\ |\psi+\phi|^2 &= |\psi|^2 + |\phi|^2 + 2|\psi||\phi|\cos(\Delta\theta) \end{aligned} \tag{1.6}$$

The quadratic form, again, reflects the fact that the size of the system is proportional to the variance of a random variable. Since the size of the system is fixed, we use unitary vectors to represent actual states. The state of the system, then, is represented by a ray in a complex inner product space.

Lastly, we need to define an expectation operator that returns the average value for each physical quantity. This operator will have to be linear under linear combination of quantities:

$$E[aX + bY|\psi] = aE[X|\psi] + bE[Y|\psi]. \tag{1.7}$$

It will not be linear under linear combination of states:

$$E[X|\psi+\phi] \neq E[X|\psi] + E[X|\phi]. \tag{1.8}$$

Yet, it will have to be proportional to the increase in size and invariant under a total change in correlation: $E[X|\tau(c)\psi] = c^*cE[X|\psi]$. This leads us to associate to each physical quantity a linear Hermitian operator X where $E[X|\psi] = \langle\psi|X|\psi\rangle$. An eigenstate ψ_0 of X corresponds to a state where all the elements of the ensemble have exactly the same value. That is, $E[(X-\bar{x})^2|\psi_0] = 0$.

Note that an inner product space can always be completed into a Hilbert space. This may, however, bring in objects that may not correspond to physical objects (i.e. infinite expectation for some quantities). In general, we believe it is better to regard the (possibly incomplete) inner product space as the physical state space and regard the completion as a mathematical device for calculation. For example, the Schwartz space seems more physically meaningful than the standard L^2 space as it gives finite expectation of all polynomials of position and momentum and, moreover, it is closed under Fourier transform.

Process with equilibria

The first type of process we consider is one with equilibria. The measurement process is recovered as a special case.

Assumption VI (Process with equilibria)**.** *Given an initial ensemble (i.e. mixed state), the final ensemble is uniquely determined and remains the same if the process is applied again.*

Under this assumption, the process can be characterized by a projection operator. Let ρ_1 be the density matrix that characterizes a mixed state. Since the final mixed state must be uniquely determined by ρ_1, it will be $\mathcal{P}(\rho_1)$ for some operator $\mathcal{P}$. Similarly, if ρ_2 is another initial mixed state, its final operator will be $\mathcal{P}(\rho_2)$. Note that, given any observable X the expectation $E[X|\rho_1] = \mathrm{tr}(X\rho_1)$ is the trace of $X\rho_1$. Similarly $E[X|\mathcal{P}(\rho_1)] = \mathrm{tr}(X\mathcal{P}(\rho_1))$.

We can always create statistical mixtures of the ensembles and we must have $E[X|a\rho_1 + b\rho_2] = aE[X|\rho_1] + bE[X|\rho_2]$ since these are classical mixtures. But since these are classical mixtures, the final state will also need to obey $E[X|a\mathcal{P}(\rho_1) + b\mathcal{P}(\rho_2)] = aE[X|\mathcal{P}(\rho_1)] + bE[X|\mathcal{P}(\rho_2)]$ for all possible X. Which means $\mathcal{P}(a\rho_1 + b\rho_2) = a\mathcal{P}(\rho_1) + b\mathcal{P}(\rho_2)$ Therefore the operator $\mathcal{P}$ is a linear operator. Moreover, the process applied twice must lead to the same result, which means $\mathcal{P}(\mathcal{P}(\rho)) = \mathcal{P}(\rho)$ for any ρ. That is, $\mathcal{P}^2 = \mathcal{P}$. Therefore $\mathcal{P}$ is a projection.

Suppose, now, that we want to measure a quantity X. We want the final outcome, the final ensemble, to be determined by the initial state, the initial ensemble. We also want the measurement to be consistent in the sense that, if it is repeated immediately after, it should yield the same result. Therefore the process will be a projection. We will also want that the process does not distort the quantity. That is, $E[X|\rho] = E[X|\mathcal{P}(\rho)]$. This means that the eigenstates of X will correspond to equilibria of the process. Moreover, subsequent measurements must give the same value, not just the same mixture. That is, if X_1 is the random variable after the first instance of the process and X_2 is the random variable after the second instance, $P(X_2 = x|X_1 = x) = 1$. This means that $E[(X - \bar{x})^2|\mathcal{P}(\rho)] = 0$ which means the eigenstates of X are the only equilibria.

The measurement process is therefore simply a special case of a process with equilibria.

Deterministic and reversible evolution

The second type of process we consider is one that is deterministic and reversible, which is the same as assumption DR.

Under this assumption, the process can be characterized by unitary evolution (i.e. the Schrodinger equation). There are multiple different ways to see this. The first relates to the more general idea that all deterministic and reversible processes must be isomorphisms in the category of states. Since the state space is an inner product space, the isomorphism is unitary evolution.

The second, is that if there is a set of quantities X_0 at time t_0 that fully identify the state (i.e. the state is the only eigenstate of those quantities), then there must be a corresponding set of quantities X_1 that fully identify the state at time t_1. This means that the evolution maps basis to basis. Moreover, given the linearity of statistical mixtures, this will also mean that a statistical distribution over X_0 will have to map to the same distribution over X_1. Therefore the evolution must map linear combinations of that basis to the same linear combination. The evolution is a linear operator. Since the total size of the irreducible system cannot change, the operator must be unitary.

The third, is by constructing a quasi-static process from processes with equilibria, much like one does in thermodynamics. The idea is that we have an infinitesimal time step, an initial state ψ_t and a final state ψ_{t+dt}. We want $P(\psi_{t+dt}|\psi_t) = 1$. This means that $|\langle\psi_{t+dt}|\psi_t\rangle|^2 = 1$. This can happen only if the difference between initial and final states is infinitesimal. That is, $\langle\psi_{t+dt}|\psi_t\rangle = 1 + \imath\epsilon dt$ where ϵ is a real number. Therefore, by convention, we can write $|\psi_{t+dt}\rangle = I + \frac{Hdt}{\imath\hbar}|\psi_t\rangle$ where H is a Hermitian operator.

Putting these perspectives together, time evolution is a unitary operator which can be written as $U = e^{\frac{H\Delta t}{\imath\hbar}}$. If we start in an eigenstate of X, that is $X|\psi_t\rangle = x_0|\psi_t\rangle$ we will end in an eigenstate $\hat{X}|\psi_{t+\Delta t}\rangle = x_0|\psi_{t+\Delta t}\rangle$ of another operator $\hat{X} = e^{\frac{H\Delta t}{\imath\hbar}} X e^{-\frac{H\Delta t}{\imath\hbar}}$.

In fact:

$$\begin{aligned} e^{\frac{H\Delta t}{\imath\hbar}} X e^{-\frac{H\Delta t}{\imath\hbar}} |\psi_{t+\Delta t}\rangle &= e^{\frac{H\Delta t}{\imath\hbar}} X e^{-\frac{H\Delta t}{\imath\hbar}} U|\psi_t\rangle \\ &= e^{\frac{H\Delta t}{\imath\hbar}} X e^{-\frac{H\Delta t}{\imath\hbar}} e^{\frac{H\Delta t}{\imath\hbar}} |\psi_t\rangle \\ &= e^{\frac{H\Delta t}{\imath\hbar}} X|\psi_t\rangle \\ &= e^{\frac{H\Delta t}{\imath\hbar}} x_0|\psi_t\rangle \\ &= x_0 U|\psi_t\rangle \\ &= x_0|\psi_{t+\Delta t}\rangle \end{aligned} \tag{1.9}$$

This is consistent with assuming there is a quasi-static process that, at every t, has equilibria identified by $e^{\frac{H(t-t_0)}{\imath\hbar}} X e^{-\frac{H(t-t_0)}{\imath\hbar}}$. Note that, unlike thermodynamics, the equilibria during the evolution are not set by external constraints but by the system itself. That is, X depends on the initial state of the system.

In this light, the measurement processes and the unitary processes can be seen as particular cases of the same type of processes, those with equilibria, which are defined as a black-box from initial to final state. This is consistent with the irreducibility assumption as the inability to describe the dynamics of the parts implicitly assumes that the dynamics of the parts is at equilibrium and sets a time-scale under which the further description of the system (i.e. non-equilibrium dynamics) would require describing the internal dynamics.

Chapter 2

Physical mathematics

This chapter presents the areas that still need to be covered to conclude the general mathematical theory of experimental science and a summary of the preliminary work done on them.

2.1 Experimental verifiability

This first part is already well developed and has been presented in chapters one to three. Possible improvements are discussed in section 2.4.

2.2 Informational granularity

The general goal of this part is to recover elements of measure theory, differential geometry, probability theory and information theory. The central theme is the ability to compare and then quantify the granularity of the description provided by different statements. The idea is to have a single unified structure which can be, in some cases, reduced to the more familiar mathematical structures.

Statement fineness

Conceptually, we want to be able to compare two statements to see which one provides a more refined description, which one provides more information. For this, we need to establish a new axiom.

Note that a theoretical domain $\bar{\mathcal{D}}$ comes with a partial order $\preccurlyeq$ that indicates whether one statement gives a **narrower**, more specific, description than the other. For example:

- *"The position of the object is between 0 and 1 meters"* $\preccurlyeq$ *"The position of the object is between 0 and 1 kilometers"*
- *"The fair die landed on 1"* $\preccurlyeq$ *"The fair die landed on 1 or 2"*
- *"The first bit is 0 and the second bit is 1"* $\preccurlyeq$ *"The first bit is 0"*

In these cases, the first statements are "contained" in the second ones, which are more general.

We need to define an additional preorder $\trianglelefteq\colon \bar{\mathcal{D}} \times \bar{\mathcal{D}} \to \mathbb{B}$ that compares two statements and tells us if the first provides a description with finer granularity than the second. Saying $\mathsf{s}_1 \trianglelefteq \mathsf{s}_2$ means that the description provided by s_1 is **finer**, gives more information, is more precise, than the description provided by s_2. For example:

- *"The position of the object is between 0 and 1 meters"* $\preccurlyeq$ *"The position of the object is between 2 and 3 kilometers"*
- *"The fair die landed on 1"* $\preccurlyeq$ *"The fair die landed on 3 or 4"*
- *"The first bit is 0 and the second bit is 1"* $\preccurlyeq$ *"The third bit is 0"*

In these cases, the first statement may not be contained or overlap with the second. The existence of this operator and its property would be an additional axiom. Fineness is a preorder, rather than an order, because it does not satisfy antisymmetry: if $\mathsf{s}_1 \preccurlyeq \mathsf{s}_2$ and $\mathsf{s}_2 \preccurlyeq \mathsf{s}_1$ then it is not necessarily true that $\mathsf{s}_1 \equiv \mathsf{s}_2$. In that case, we will say that the two statements are **equigranular**, noted $\mathsf{s}_1 \doteq \mathsf{s}_2$.

Note how statements about geometry, probability and information all satisfy the same concept. In fact, each of these structures will generate a preorder on the statements. The general question is what are the necessary and sufficient conditions on the preorder to be able to recover those structures.

Measure theory

Conceptually, a measure allows one to assign a size to a set. For us, a theoretical set is really a statement, so we want to assign sizes to statements that represent the coarseness of the description they provide.

The construction should, roughly, proceed as follows. Let $\bar{\mathcal{D}}_X$ be a theoretical domain. We select a unit statement $\mathsf{u} \in \bar{\mathcal{D}}_X$. We define, in some way, the set $\bar{\mathcal{D}}_\mathsf{u} \subseteq \bar{\mathcal{D}}_X$ which contains all statements that are comparable to u. We then try and construct a measure $\mu_\mathsf{u} : \bar{\mathcal{D}}_\mathsf{u} \to \mathbb{R}$ such that $\mu_\mathsf{u}(\mathsf{u}) = 1$. By a measure, we mean that μ_u is additive over incompatible statements (i.e. disjoint sets of possibilities). That is, if $\mathsf{s}_1 \nparallel \mathsf{s}_2$, we have $\mu_\mathsf{u}(\mathsf{s}_1 \vee \mathsf{s}_2) = \mu_\mathsf{u}(\mathsf{s}_1) + \mu_\mathsf{u}(\mathsf{s}_2)$. We want the measure to respect the fineness preorder, to be monotonic. That is, if $\mathsf{s}_1 \preccurlyeq \mathsf{s}_2$ then $\mu_\mathsf{u}(\mathsf{s}_1) \leq \mu_\mathsf{u}(\mathsf{s}_2)$.

Originally, we thought that these measures would have to be always additive and therefore we starting adding suitable axioms on fineness. However, we realized that, in the context of quantum mechanics, the measure cannot be additive if it has to agree with the von Neumann entropy. Worse, it is not even monotonic (i.e. a broader statement is not necessarily coarser). More conceptual work needs to be done to understand the issue.

Note that we have essentially one measure for each equivalence class defined by fineness. This is intended. One reason a single measure is not sufficient for our work is because we need to compare statements of "different infinities". If we have a single measure, we can only compare objects with a finite measure. All objects with zero measure (or infinite measure) are indistinguishable. For example, we want to say:

- s_1 = *"The horizontal position of the object is exactly 0 meters"*
- s_2 = *"The horizontal position of the object is exactly 1 or 2 meters"*
- s_3 = *"The horizontal position of the object is between 0.5 and 1.5 meters"*
- s_4 = *"The horizontal position of the object is between 1.5 and 3.5 meters"*
- $\mathsf{s}_1 \preccurlyeq \mathsf{s}_2 \preccurlyeq \mathsf{s}_3 \preccurlyeq \mathsf{s}_4$
- $\mathsf{s}_1 \not\succcurlyeq \mathsf{s}_2 \not\succcurlyeq \mathsf{s}_3 \not\succcurlyeq \mathsf{s}_4$

Fineness may also capture the concept of physical dimension. In fact, two descriptions in the same units are "finitely comparable" in the sense that one gives a finer description than the other by a finite factor. Descriptions of different units are either "infinitely comparable"

(e.g. areas are always bigger than lengths) or not comparable (e.g. position and momentum). Consider a two dimensional phase space of a classical system. Points should be comparable and in fact should be equigranular $\doteq$ so that we can compare sets of finitely many points. Areas are also comparable to each other, and are comparable to points (i.e. they are infinitely bigger). However, vertical lines (i.e. ranges in momentum alone) are not comparable to horizontal lines (i.e. ranges in position alone). Symplectic geometry, in fact, gives a size to areas and not to lines. Mathematically, this should be clarified when one is trying to define the domain of the measure $\bar{\mathcal{D}}_{\mathsf{u}}$.

Probability

Conceptually, probability is recovered as a measure restricted to a particular subset. The idea is that you take two statements, such as *"the die landed on 2"* given that *"the die has 6 sides and it is fair"*, and you ask what fraction of the possibilities compatible with the second is also compatible with the first. This defines the conditional probability.

Let $\mathsf{s}_1, \mathsf{s}_2 \in \bar{\mathcal{D}}$ be two theoretical statements. Then the probability of s_2 given s_1 is

$$P(\mathsf{s}_2|\mathsf{s}_1) = \mu_{\mathsf{s}_1}(\mathsf{s}_1 \wedge \mathsf{s}_2) = \frac{\mu_{\mathsf{u}}(\mathsf{s}_1 \wedge \mathsf{s}_2)}{\mu_{\mathsf{u}}(\mathsf{s}_1)} \tag{2.1}$$

which quantifies the fraction of possibilities compatible with s_1 that are also compatible with s_2.

If we take the certainty $\top$ as a unit, we have a probability measure for the whole space. However, since we can take different statements as a unit, we will be able to distinguish between the following cases:

- $P($ *"n is odd"* $|$ *"n is picked fairly from all integers"*$) = 1/2$
- $P($ *"n is between 0 and 9"* $|$ *"n is picked fairly from all integers"*$) = 0$
- $P($ *"n is 3"* $|$ *"n is picked fairly from all integers"*$) = 0$
- $P($ *"n is 3"* $|$ *"n is between 0 and 9"* $\wedge$ *"n is picked fairly from all integers"*$) = 1/10$

Differentiability

We want to construct a notion of differentials and differentiability that is the same for all spaces, even infinite dimensional ones. When introducing derivatives, this is typically done by taking limits of differences, and therefore differentiability is the existence of those limits. In differential topology, this notion is used to define differentiability of manifolds in terms of differentiability of coordinates, and then differentials are defined as linear functions of vectors. That is, the differentials defined on the coordinates of a particular chart are technically not the same objects as the differentials defined on the space.

The idea is to define differentiability on the vector space structure alone. That is, given two vector spaces V and W, a map $f : V \to W$ is differentiable if it becomes linear in the neighborhood. We would first define a differential as a sequence of vectors $\{v_i\}_{i=1}^{\infty} \in V$ such that there exists a vector $t \in V$ and a sequence of non-zero elements $\{a_i\}_{i=1}^{\infty} \in \mathbb{R}$ that converges to 0 for which

$$\lim_{i \to \infty} \frac{v_i}{a_i} = t.$$

We call t the **tangent vector** of the differential and $\{a_i\}_{i=1}^{\infty}$ its **convergence envelope**. Note that, given a sequence v_i, these are not unique. We note $dv[a_i\, t]$ the differential with its

tangent vector and convergence envelope. One can show that every differential can be written as $v_i = a_i t_i$ where t_i converges to t.

We can now study how a map $f : V \to W$ maps differentials. Given a sequence $\{v_i\}_{i=1}^{\infty} \in V$, we can define $w_i = f(v_i)$. If, additionally, we have a differential $dv[a_i\, t]$, we can define the sequence $\{w_i\}_{i=1}^{\infty} = \{f(v_i + a_i t_i) - f(v_i)\}_{i=1}^{\infty}$. Now, the observation here is that if the map is linear, the sequence $\{w_i\}_{i=1}^{\infty}$ will be a differential with tangent vector $f(t)$ and convergence envelope a_i. But any map that is locally linear will have the same property, given that differentials are local objects. Therefore we say f is differentiable at $v \in V$ if there exists a map $d_v f|_{v_0} : V \to W$ such that $\{w_i\}_{i=1}^{\infty} = dw[a_i\, d_v f|_{v_0}(t)]$.

From a preliminary study, this would work on any vector space, regardless of dimension or field (i.e. real, complex, rational, ...).

Differential geometry/geometric measure theory

In the reverse physics chapter about classical mechanics we have seen that forms can be understood as modeling additive functionals of subregions. We need to connect those ideas to the rest of the formal framework.

Conceptually, we want to assign quantities to regions instead of points. If we assume these quantities are additive, the idea is that we can decompose them into the sum of infinitesimal contributions at each point. Therefore the differential objects exist as the limit of infinitesimal decomposition. This, again, reflects the overall spirit of the project that compels us to start from physically well defined entities (in this case the quantities associated with finite regions) and derive the theoretical ones (in this case the infinitesimal contributions that are integrated).

Let $\bar{\mathcal{D}}_X$ be a theoretical domain and $U \in \Sigma_X$ a theoretical set. This represents the region associated to our measurement. Let $\bar{\mathcal{D}}_Y$ be a theoretical domain and $R \in \Sigma_Y$ a theoretical set. This represents the possible values found. Our starting point consists of statements like:

- *"the amount of mass inside volume U is within range R"*
- *"the force applied to surface U is within range R"*
- *"the energy used to move the object along the line U is within range R"*

These are finite precision statements of a quantity associated to a region of finite size.

The first step is to group statements within the same region U into subdomains $\bar{\mathcal{D}}_{U \to Y}$. We can then show how the possibilities for each $\bar{\mathcal{D}}_{U \to Y}$ reduce to statements like:

- *"the amount of mass inside volume U is precisely y"*
- *"the force applied to surface U is precisely y"*
- *"the energy used to move the object along the line U is precisely y"*

These are infinite precision statements of a quantity associated to a region of finite size. We define $S \subseteq \Sigma_X$ as the type of region (i.e. volumes vs surfaces vs lines) upon which the functional is defined and therefore we have a functional $f : S \to Y$ which tells us the exact value of the quantity in each region.

Then we study the case where f is a real linear k-functional, meaning:

- the possibilities X are identified by a set of real values; that is, X with the natural topology is a manifold
- the domain is all k-dimensional surfaces S^k; that is, the submanifolds of dimension k

- the co-domain is the reals; so we have $f : S^k \to \mathbb{R}$
- the functional is additive over disjoint sets; that is, $F(U_1 \cup U_2) = F(U_1) + F(U_2)$ if $U_1 \cap U_2 = \varnothing$
- the functional commutes with the limit; that is, $\lim_{i\to\infty} F(U_i) = F(\lim_{i\to\infty} U_i)$

Under these conditions (and possibly others) one can express the functional as a sum of infinitesimal contributions. That is, $f(U) = \int_U \omega(dU)$, where ω represents a suitable k-form.

Note that there is not a unique way to perform this decomposition. For example, if $f(U)$ is the total mass in the volume, $\omega(dU)$ is the density in the infinitesimal volume. If we change the density at a single point, the integral does not change and only the integral is physical. These are the types of issues that still need to be solved.

Stokes' theorem and exterior derivatives. One interesting application of this viewpoint is that we can understand things like Stokes' theorem, exterior derivative and the difference between closed and exact forms directly on the finite functionals.

Let $\partial : S^k \to S^{k-1}$ be the boundary operator that, given a surface σ^k, returns the boundary $\partial\sigma^k$ which is of dimension $k-1$. We have $\partial\partial\sigma = \varnothing$ for any surface of any dimensionality.

Let F_k be the space of linear k-functionals. We can define the boundary functional operator $\partial : F_k \to F_{k+1}$ such that $\partial f(\sigma) = f(\partial\sigma)$. That is, given a functional that acts on k-surfaces we can always construct one that acts on $k+1$-surfaces by taking the boundary of the $k+1$-surface and giving it to the first functional. Note that $\partial\partial f(\sigma) = \partial f(\partial\sigma) = f(\partial\partial\sigma) = f(\varnothing) = 0$, so the boundary functional of the boundary functional is the null functional, the one that returns zero for every k-surface. What we should be able to prove is that if ω is the k-form associated with f, $d\omega$ is the $k+1$-form associated with ∂f. In other words, Stokes' theorem essentially becomes a definition of the boundary functional and the calculation of the expression for $d\omega$.

We say a surface is contractible if it can be reduced to a point with a continuous transformation. A functional is closed if it is zero for all closed contractible surfaces. It is exact if it is zero for all closed surfaces. All boundary functionals are exact since $\partial f(\partial\sigma) = f(\partial\partial\sigma) = 0$. The form associated to a closed functional will be closed while the form associated to an exact functional will be exact.

2.3 States and processes

The general goal of this part is to give general definitions of states and processes that are always valid and are captured by a fundamental mathematical framework. Different theories would then specialize these basic definitions for different circumstances.

Processes

A process is an experimental domain $\mathcal{P}$ that contains all the possible statements of the systems under study for all possible times. We call evolutions the possibilities E of the domain, as they represent the complete description of all systems at all times.

We define a time parameter $t \in T \subseteq \mathbb{R}$. We group all statements relative to a system of interest at a particular time into a time domain $\mathcal{D}_t$. We call snapshots the possibilities X_t of each time domain. A possible trajectory is a sequence $\{x_t\}_{t\in T}$ such that $x_t \in X_t$ for all $t \in T$ and $e \preccurlyeq \bigwedge_{t\in T} x_t$ for some $e \in E$. That is, there is an evolution for which the system will be described by that sequence of snapshots.

A process is deterministic if for all possible trajectories $x_{t_0} \preccurlyeq x_{t_1}$ for all $t_0 \leq t_1$. A process is reversible if for all possible trajectories $x_{t_1} \preccurlyeq x_{t_0}$ for all $t_0 \leq t_1$. Recall that narrowness between the possibilities of two domains means there is an experimental relationship. Therefore, if the process is deterministic, we can write a causal relationship $f : X_{t_0} \to X_{t_1}$ such that $x_0 \preccurlyeq f(x_0)$.

Once we derive a measure $\mu_{\mathsf{u}} : \bar{\mathcal{P}}_{\mathsf{u}} \to \mathbb{R}$, we can define the evolution entropy as $\log \mu_{\mathsf{u}}$. As the measure is multiplicative for independent systems, the evolution entropy will be additive making it an extensive property. The evolution entropy of a system at a time is defined to be the evolution entropy $\log \mu_{\mathsf{u}}(x_t)$ of the snapshot at that time. Under a deterministic process, the evolution entropy can never decrease: $\log \mu_{\mathsf{u}}(x_{t_0}) \leq \log \mu_{\mathsf{u}}(x_{t_1})$ since $x_{t_0} \preccurlyeq x_{t_1}$ for all $t_0 \leq t_1$ and therefore $\mu_{\mathsf{u}}(x_{t_0}) \leq \mu_{\mathsf{u}}(x_{t_1})$. If the process is also reversible, then $\log \mu_{\mathsf{u}}(x_{t_0}) = \log \mu_{\mathsf{u}}(x_{t_1})$.

These definitions give a very general setting to describe a process and already find a quantity that cannot decrease during deterministic evolution.

States

Conceptually, states represent description of the system, and only of the system, regardless of time. Therefore the state space is not a set of statements, but a "template" for a set of statements that can be "instantiated" at different times.

The idea is that a state space $\mathcal{S}$ comes equipped with a function $\iota : \mathcal{S} \times T \to \bar{\mathcal{P}}$ such that $\iota(\mathcal{S}, t) = \bar{\mathcal{D}}_t$. That is, it maps the state space and its statements to the particular time domain that represents the system at that particular time. Specifically, states of the system will be mapped to snapshots of the system.

The structure of the state space will not be, in general, isomorphic to each particular time domain. In a particular process at a particular time some states may not be accessible, so some states will be mapped to an impossibility. Or there may be correlations with other system, so the snapshot will provide more information (will be narrower) than the states themselves.

The relationships defined on the state space will be equivalent to the ones in the time domain if and only if the time domain of the system is independent from the time domain of the other systems. In other words: the state space represents the system and its properties when the system is independent. This also means that, to be able to define a system, we need to have a process that renders it independent from other systems.

When the system is independent from all others, the description is coarser than in the case of when there are correlations. Note that to a coarser description is associated a higher process entropy. Processes that render the system independent are exactly the ones that maximize the process entropy. We can associate a state entropy to each state, which is the process entropy associated to that description when the system is independent.

While it is still not clear what can be derived and what must be imposed, the overall goal is to understand what assumptions are needed to construct state spaces. One result should be that processes that isolate the system are implicitly needed, which forms the basis of requiring entropy maximization. All states are therefore equilibria of those processes (i.e. symmetries of the group of processes). Conceptually, this maps well with all branches of physics as all state spaces come equipped with some structure which, in the end, is connected to entropy/probability/measure.

2.4 Open questions and possible extensions

Here we note some thoughts and ideas about open problems and possible extensions to the general theory.

Homogeneity of an experimental domain

It may be interesting to characterize some notion of homogeneity that makes all possibilities in a domain "equally verifiable", that no possibility is "special" compared to the others in terms of experimental verifiability. For example:

- the "extra-terrestrial life" domain is not homogeneous because one possibility can be verified while the other cannot
- the integers and reals are the only linearly ordered quantities where all contingent statements are the same experimentally: all decidable and none decidable
- phase transitions are special, as knowing whether a system is in a mixed state is decidable, so a domain with phase transitions is not homogeneous

It is not clear how this notion should be implemented and how exactly it would be useful. It may give a reason to expect a complete domain (the residual possibility is the only one that is not compatible with any contingent verifiable statement, so the domain would not be homogeneous) and also that all possibilities are approximately verifiable (if one is able to prove that, in any domain, at least one possibility is approximately verifiable).

Predictive relationships

Another way to characterize relationships between domains could be in terms of predictions, what statements of one domain can tell about the other. That is, we give a theoretical statement on one and look for the best prediction (i.e. narrowest theoretical statement broader than the original) for the other.

For example, if a domain is independent from another, any theoretical statement should predict the certainty on the other. If a domain is dependent on another, any theoretical statement should predict an equivalent statement.

A possible approach. Let $\mathcal{D}_X$ and $\mathcal{D}_Y$ be two experimental domains and $\bar{\mathcal{D}}_X$ and $\bar{\mathcal{D}}_Y$ their respective theoretical domains. Now we construct the function $\pi : \bar{\mathcal{D}}_X \to \bar{\mathcal{D}}_Y$ such that given $\mathsf{s}_X \in \bar{\mathcal{D}}_x$ and $\mathsf{s}_Y \in \bar{\mathcal{D}}_Y$ such that $\mathsf{s}_X \preccurlyeq \mathsf{s}_Y$ we always have $\mathsf{s}_X \preccurlyeq \pi(\mathsf{s}_X) \preccurlyeq \mathsf{s}_Y$. In other words, it should map to the narrowest broader statement in $\bar{\mathcal{D}}_Y$.

In principle, we can even extend $\pi : \mathcal{S} \to \bar{\mathcal{D}}_Y$ to be defined on the whole context. In that case, π can be proven to be a projection. This map should be able to characterize the relationship between domains. For example, if $\pi(\bar{\mathcal{D}}_X) = \{\top, \bot\}$ then the domains should be independent. If $\pi(\bar{\mathcal{D}}_X) = \bar{\mathcal{D}}_Y$ the domains should be dependent.

Defining structures on experimental domains

Some mathematical structures are defined on points (i.e. vector spaces, ordering) and others on their σ-algebras. In our context, verifiable statements are the only elements that are actually physical, therefore it would be nice to always define the structures on the experimental domain (i.e. the topology) and show that it induces a unique structure on the theoretical statements and possibilities (and vice-versa).

We have already implemented this approach in a couple of areas. Theoretical domains are constructed from experimental domains (see **??**) and so are the possibilities (see **??**). Theorem **??** shows that causal relationship on the possibilities is equivalent to an inference relationship on the verifiable domain. Theorem **??** shows that ordering of the possibilities is equivalent to the ordering of the basis according to narrowness.

We need to understand how this can be achieved for other structures, such as measures, metrics, groups, vector spaces, inner products, ...

Space of possible combined domains

It should be possible to better characterize the space of all possible combined domains. As we show in **??** that the space of the possible experimental relationships is the space of topologically continuous functions, there should be an analogue for the space of all possible combined domains. For example, one should be able to show that the combined domain is an immersion within the product topology. Is that the only constraint? How can that be characterized? Can we create an experimental domain to distinguish them?

Limited precision

One area we could explore for new physical ideas is what happens if we assume that the precision cannot be arbitrarily decreased. How is it different from the continuous case? Here are some preliminary ideas.

The limited precision case cannot simply lead to a discrete topology. The standard topology of the reals is not the limit of the integer topology since it is not discrete. Most likely, the limited precision case will need to have uncountable possibilities so that the limit to arbitrary precision can work well.

The main cause of confusion is that, in the continuous case, whether the precision of two statements overlap determines whether the statements are compatible. For example, "the position is between 0 and 1 meters" and "the position is between 2 and 3 meters" are both incompatible and not overlapping. This cannot be the case for limited precision. The possibilities themselves must be incompatible with each other but some of them must overlap, or we would simply have a discrete topology. That is, suppose that 1 unit is the precision limit, the statements "the position is between 0 and 1" and "the position is between 0.5 and 1.5" are incompatible because if we verify one we cannot verify the other. If we could verify them both, we would measure at a smaller precision. So, overlapping cannot be defined in terms of incompatibility.

Whether two statements overlap cannot be determined through incompatibility but must be recovered from the precision of the disjunction. Suppose we have the following arbitrary precision statements.

s_1 = *"the position is between 0 and 1 meter"*
s_2 = *"the position is between 0.5 and 1.5 meter"*
s_3 = *"the position is between 2 and 3 meter"*

The precision associated to all statements will be one meter. The precision for $\mathsf{s}_1 \vee \mathsf{s}_2$ will be one meter and a half while the precision for $\mathsf{s}_1 \vee \mathsf{s}_3$ will be two meters. That is: the precision for non overlapping statements sums. It may even be the case that if the precision sums, statements must be incompatible but the converse is what fails.

This means that the measure we put on the possibilities cannot represent the precision anymore. That is, $d\mu \neq dx$. We can imagine a relationship like $dx^2 = d\mu^2 + 1$. This would both make the precision go to 1 when the measure goes to 0 and $dx \simeq d\mu$ for large μ.

Chapter 3

Quantum mechanics

3.1 The postulates of quantum mechanics

In this section we will review the standard formulations of quantum mechanics in terms of projective Hilbert spaces.

Proposed nomenclature

- State vector $|\psi\rangle \in \mathcal{H}$
- ? Wave function $\psi(x)$ $\psi(s_z)$
- State $\psi \in P(\mathcal{H})$
- Ensemble $\rho \in D(\mathcal{H})$
- ? Space of observables, space of bounded operators, space of unitaries, ...

Sketch of results

How to break up the different assumptions for quantum states.

QP Space of pure states by themselves tells us nothing. Complex projective spaces are symplectic spaces, and are symplectic manifolds in finite dimension. We can write Hamiltonians and do classical mechanics with probability distributions over the states.

QE Convex space of density operators tells you more, but not everything. We know they form a convex subset of a vector space. We know which ensembles allow multiple decompositions and we can define observables as statistical quantities (i.e. linear functionals of the ensembles). However, we do not know what the entropy is. For example, take the Bloch ball with the standard entropoy. Take an inner ball and discard the rest. Now, the entropy difference between pure states and maximally mixed state is not 1, therefore a variable cannot extract one bit of information. This means that opposite states are not orthogonal (i.e. you could mistake up for down with a single measurement).

BR Conjecture. If we add orthogonal decomposability, we should be able to recover the Born rule, and the Born rule tells us ensembles are orthogonally decomposable. One should be able to show that all "effects" (i.e. linear functionals of ensembles that are either zero or one in the edges orthogonal to the gradients) have orthogonal eigenspaces of either zero or one, and therefore can extract one bit of information. That should be enough to define the entropy, find the Born rule, etc...

Define BT (Base Theory) as a set of ensembles that can be mixed.
Over BT

? QP is not enough to identify classical vs quantum
? QE $\implies$ QP
? QE is enough to rule out classical, but not to fully identify quantum
? QE + BR = QST

The following are equivalent over BT

QP-RAY A pure quantum state is represented by a ray
QE-ENS Ensembles are density operators (recover pure states as extreme points, recover ensembles as mixtures)
QP-SUBS A pure state is a density operator with only one non-zero eigenvector
QE $\implies$ QP Show that pure states are ensembles with only one non-zero eigenvector (i.e. they are extreme points)
QP-PROJ A pure state is a projector
QE-OBS The set of random variables over probability distributions that are mapped univocally under multiple decompositions fully identify the space of quantum ensembles (i.e. we can either start with the space of quantum ensembles and rule out those random variables that are not the same over equivalence classes of probability distributions, or we can start with the "nice" random variables, and find the ensembles as equivalence classes of probability distributions).

The following are established over QE

? Superposition is multiple decomposition
? The linearity of statistical mixing is a physical requirement. The linearity of the Hilbert space is a mathematical convenience.
? The choice of a basis in the Hilbert space is a choice of a maximal set of orthogonal states (not necessarily mutually exclusive) plus a choice of gauge (i.e. a phase for each element of the basis).

The following are equivalent over QE

BR-ME Orthogonal states are mutually exclusive
BR-PROB The probability of measuring a final state given an initial state is well-defined.
BR-ANG The born rule returns the cosine(?) of the angles between pure
BR-ENT The entropy is the von Neumann entropy

Over BT

QST QE + BR

The following are equivalent over QST

DR-SCEQ The equation
DR-UNIT Unitary evolution
DR-INN Preservation of inner product
DR-NORM Preservation of norm
DR-UBOR Square of inner product infinitesimally close
DR-PERP Perpendicular change

DR-OBAS Preserves orthogonal basis
DR-EV Deterministic and reversible evolution
DR-PROB Probability distribution preserved
DR-INFO The evolution preserves information entropy

Over QST+PM $\implies$ DR-PSEQ and DR-MSEQ, and QST+DR $\implies$ PM-DEQ

DR-PSEQ The evolution is quasi-static process
DR-MSEQ The evolution is an infinite sequence of reversible measurements.
PM-DEQ Projection measurements are process with equilibria that commute with a det/rev process.

Random things

- tensor product/composite systems
- Fermion and Bosons as indistinguishibility of components (requires tensor product)

In QST, infinite dimension

- A pure state is in the domain of an observable if and only if the expectation of the second moment is finite

States

The first postulate of quantum mechanics characterizes how states are represented in the mathematical framework.

Postulate ST (State postulate)**.** *The state space of a quantum system is represented by the projective space* $\mathrm{P}(\mathcal{H})$ *of a complex Hilbert space* $\mathcal{H}$*. A state is represented by a ray of the Hilbert space* $\mathcal{H}$*. An ensemble is represented by a density operator* $\rho : \mathcal{H} \to \mathcal{H}$*, that is a positive semi-definite trace one self-adjoint operator.*

In terms of notation, we will mostly follow the convention most used by physicists,[1] with some addition or modification to make the connection to the math more clear. Mathematically, a complex Hilbert space $\mathcal{H}$ has three defining properties. It is a vector space, meaning that it is closed under addition $+ : \mathcal{H} \times \mathcal{H} \to \mathcal{H}$ and scalar multiplication $\cdot : \mathbb{C} \times \mathcal{H} \to \mathcal{H}$. It has an inner product, meaning that it is equipped with a map $\langle \cdot | \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ that is positive-definite, linear in the second argument and conjugate symmetric. It is a complete metric space with respect to the distance function induced by the inner product, meaning that, given a sequence of vectors $|\psi_i\rangle$, if the series given by their norm $\sum_i |\psi_i|$ converges, then the series of given by the vectors $\sum_i |\psi_i\rangle$ converges as well. A ray of a Hilbert space is a one dimensional subspace. That is, given a vector $|\psi\rangle$, the corresponding ray is the set of all vectors given by $c|\psi\rangle$. With a slight abuse of notation, we will use ψ to indicate the ray identified by the vector $|\psi\rangle$.

The second postulate of quantum mechanics characterizes observables.

Postulate OBS (Observable postulate)**.** *An observable is represented by a self-adjoint operator* $O : \mathcal{H} \to \mathcal{H}$*. Given a state* $|\psi\rangle$*, the expectation is given by* $\langle \psi | O | \psi \rangle$*, and, given an ensemble* ρ*, the expectation is given by* $\mathrm{tr}(O\rho)$*.*

[1] Vectors will be noted as $|\psi\rangle$ instead by a simple letter v. The inner product will be noted as $\langle \phi | \psi \rangle$, with a vertical line instead of a comma, and it will be linear in the first term instead of the second term.

Postulate PROJ (Projection measurement postulate)**.** *A projection measurement is defined by a set* $M = \{\mathbf{1}_i\}$ *of projectors, representing the measurement outcomes, such that* $\sum_i \mathbf{1}_i = I$*. If* $|\psi\rangle \in \mathcal{H}$ *represents the state before the measurement interaction, the ensemble after the interaction will be* $\rho_M = \sum_i \mathbf{1}_i \mathbf{1}_\psi$ *and the state after the measurement of the i-th outcome is* $|\phi_i\rangle = \frac{\mathbf{1}_i|\psi\rangle}{\sqrt{\langle\psi|\mathbf{1}_i|\psi\rangle}}$ *with probability* $\langle\psi|\mathbf{1}_i|\psi\rangle$*. If* ρ *is the ensemble before the measurement interaction, the ensemble after the interaction will be* $\rho_M = \sum_i \mathbf{1}_i \rho$ *and the state after the measurement of the i-th outcome is* $\hat{\rho}_i = \frac{\mathbf{1}_i \rho}{\mathrm{tr}(\mathbf{1}_i \rho)}$ *with probability* $\mathrm{tr}(\mathbf{1}_i \rho)$*.*

Postulate UNIT (Time evolution postulate)**.** *The time evolution of a state vector* $|\psi\rangle$ *is given by* $d_t|\psi(t)\rangle = \frac{H}{i\hbar}|\psi(t)\rangle$ *where* H *is a self-adjoint operator. The time evolution of an ensemble* ρ *is given by* $d_t\rho = \left[\frac{H}{i\hbar}, \rho\right]$*.*

Postulate COMP (Composition postulate)**.** *Given two quantum systems associated with the Hilbert spaces* $\mathcal{H}_1$ *and* $\mathcal{H}_2$*, their composite system is associates with the tensor product* $\mathcal{H}_1 \otimes \mathcal{H}_2$*.*

Review of the mathematical formulation of quantum mechanics * wave-function (postulates - rays of Hilbert space) - states are rays in Hilbert space - observables are Hermitian operator inner product, Born rule, projection as final state - composite systems (tensor product) - unitary evolution * mixtures and entropy

Composite system * Show that system composition of quantum systems that is itself a quantum system gives the tensor product * Review the arguments in the paper * see if using a Segre embedding would make our life easier

Bloch sphere review

The state space of a two-state quantum system (i.e. qubit) is of fundamental importance as it allows us to visualize and therefore better understand almost all features of quantum mechanics. The space of pure states is isomorphic to a unit 2-sphere, with opposite points corresponding to orthogonal states, meaning that we can understand each pure state as a unit vector along a direction of three dimensional space. Since this is also the space of a spin 1/2 system, we will often label states as spatial directions. Let us quickly review the main features.

Let $|z^+\rangle$ and $|z^-\rangle$ be two orthonormal states, representing, for example, the two possible directions of spin along the vertical direction. Every other normalized state can be expressed as $\rho|z^+\rangle + \sqrt{1-\rho^2}e^{\imath\varphi}|z^-\rangle$ with $\rho \in [0,1]$ and $\varphi \in [0, 2\pi)$. Normalization, in fact, imposes that the square of the norm of the coefficient sum to one. The arbitrariness of the phase allows us to impose the first coefficient to be real. Note that, for any $p \in (0,1)$, we can choose an arbitrary φ, meaning that we are choosing a point on a circle. At the endpoints $p \in \{0,1\}$, however, there is no longer a choice of φ. The topology, then, is the one of the sphere. We can therefore draw a sphere with $|z^+\rangle$ and $|z^-\rangle$ respectively at the top and at the bottom. The choice of ρ will select a horizontal plane, whose intersection with the sphere will be a circle. The choice of φ will pick a point on that circle. That is, ρ and φ can be understood as a sort of "cylindrical coordinates over a sphere." We can also choose spherical coordinates (θ, φ) such that $\rho = \cos\theta$.

TODO: show that ensemble parametrization in spherical coordinates is not affine (meaning the mixture of the ensembles does not give the mixture of the coordinates)

More remarkably, the interior points of the sphere represent mixed states. Therefore the whole space, the Bloch ball, represents the full space of ensembles for a two-state quantum systems. Given two ensembles ρ_1 and ρ_2, the statistical mixture $\rho = p\rho_1 + (1-p)\rho_2$ will be the point in between the two. The coefficient for each endpoint is higher the closer the mixture is to that point, meaning it is proportional to the distance to the other endpoint. That is,

$$\rho = \frac{\overline{\rho\rho_2}}{\overline{\rho_1\rho_2}}\rho_1 + \frac{\overline{\rho_1\rho}}{\overline{\rho_1\rho_2}}\rho_2. \tag{3.1}$$

3.2 States and ensembles

As we have already seen in classical mechanics, re-expressing the same mathematical framework in different equivalent ways brings out new understanding. This will be even more clear in quantum mechanics, as the standard Hilbert space formulation is as good for calculations it is bad for understanding.

Different Representations of pure states

First of all, we should note that wave-functions do not provide a unique representation for states. The reader may already be aware that a change in norm or phase does not change the physics. Intuitively, all the physics is given by the born rule

$$p(\phi|\psi) = \frac{\langle\phi|\psi\rangle\langle\psi|\phi\rangle}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle}. \tag{3.2}$$

If we chage ψ with $\rho e^{\imath\theta}\psi$, we have

$$p(\phi|\rho e^{\imath\theta}\psi) = \frac{\langle\phi|\rho e^{\imath\theta}\psi\rangle\langle\rho e^{\imath\theta}\psi|\phi\rangle}{\langle\rho e^{\imath\theta}\psi|\rho e^{\imath\theta}\psi\rangle\langle\phi|\phi\rangle} = \frac{\rho e^{\imath\theta}\rho e^{-\imath\theta}}{\rho e^{-\imath\theta}\rho e^{\imath\theta}}\frac{\langle\phi|\psi\rangle\langle\psi|\phi\rangle}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle} = \frac{\langle\phi|\psi\rangle\langle\psi|\phi\rangle}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle}. \tag{3.3}$$

There is another detail, however, that is often neglected, maybe because it applies only to continuous variables. The inner production between two wave functions is given by:

$$\langle\phi|\psi\rangle = \int_X \phi^*(x)\psi(x)dx. \tag{3.4}$$

Note that integrals do not change if we change the value of the wave-function over a set of measure zero. For example, if we changed $\psi(x)$ such had it returned zero over the rationals, the inner product wouldn't change. If the physics is given by the Born rule through the inner product, then the physics does not change. Mathematically, when we write the vector $|\psi\rangle$ we are actually talking about the set of all functions that are equivalent to the wave function $\psi(x)$. This is not something peculiar to quantum mechanics. In fact, this happens in classical probability as well. If we have a probability density $\rho(x,p)$, a change over a measure zero set will not change the expectation of any variable.

Since the representation in terms of Hilbert spaces already removes the second issue, is there a representation of states that eliminates the first as well? Turns out that there is. In a sense, working with Hilbert space is somewhat a conceptual no man's land. We are aware that we are dealing with probabilistic objects, but we are not really working with full fledged statistical mixtures. A lot things become clearer once we look at the quantum statistical framework as a whole and compare it to the classical statistical framework.

Let us look, then, at the space of all possible statistical mixtures. In quantum statistical mechanics, a state is represented by a density operator ρ. Mathematically, this is an operator $\rho : \mathcal{H} \to \mathcal{H}$ that is positive semi-definite trace one self-adjoint operator. In fact, since $\langle\psi|\rho|\psi\rangle$ returns the probability to measure ψ given the state ρ, ρ is an observable. Since the probability cannot be negative, ρ is positive semi-definite. Since the probability has to sum to one, it is a trace one operator.

Every pure state ψ can be represented as the density operator $\rho_\psi = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle}$. Note that, in fact, that ρ_ψ is trace one and a change of phase will leave it unchanged. Also note that $\frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle} = \mathbf{1}_\psi$ is the projector corresponding by the subspace spanned by ψ.

Insight 3.5. *A state is equally represented by a ray $c|\psi\rangle$, by a density operator ρ_ψ where $c|\psi\rangle$ is the only non-zero eigenvector, or by a the projector $\mathbf{1}_\psi$ corresponding to the subspace $c|\psi\rangle$.*

We now have three equivalent ways to represent pure states.

Two linearities: superposition and statistical mixing

A linear combination, a superposition, $\sum c_i|\psi_i\rangle$ will give us another state, but is somewhat unclear what this operation means. First of all, we have the problem that the superposition is physically unique up to a total phase, so only the phase differences are physically significant. Moreover, the meaning of the norm of the coefficient c_i is also ill defined. The overall intuition one gets is that $|c_i|^2$ corresponds to probability, but this does not actually work. In fact, we have

$$\begin{aligned}
|\phi\rangle &= c_1|\psi_1\rangle + c_2|\psi_2\rangle \\
\langle\phi|\phi\rangle &= |c_1|^2\langle\psi_1|\psi_1\rangle + c_1^* c_2\langle\psi_1|\psi_2\rangle + c_2^* c_1\langle\psi_2|\psi_1\rangle + |c_2|^2\langle\psi_2|\psi_2\rangle \\
\langle\psi_1|\phi\rangle &= c_1\langle\psi_1|\psi_1\rangle + c_2\langle\psi_1|\psi_2\rangle \\
p(\phi|\psi_1) &= \frac{\langle\phi|\psi_1\rangle\langle\psi_1|\phi\rangle}{\langle\psi_1|\psi_1\rangle\langle\phi|\phi\rangle} \\
&= \frac{(c_1^*\langle\psi_1|\psi_1\rangle + c_2^*\langle\psi_2|\psi_1\rangle)(c_1\langle\psi_1|\psi_1\rangle + c_2\langle\psi_1|\psi_2\rangle)}{\langle\psi_1|\psi_1\rangle(|c_1|^2\langle\psi_1|\psi_1\rangle + c_1^* c_2\langle\psi_1|\psi_2\rangle + c_2^* c_1\langle\psi_2|\psi_1\rangle + |c_2|^2\langle\psi_2|\psi_2\rangle)}.
\end{aligned} \tag{3.6}$$

In the special case where ψ_1 and ψ_2 are orthonormal, the above simplifies to

$$p(\phi|\psi_1) = \frac{|c_1|^2}{|c_1|^2 + |c_2|^2} \tag{3.7}$$

and therefore, if $|c_1|^2 + |c_2|^2 = 1$, we recover the probability. But this is a special case which does not work in general. And building our understanding on special cases that do not work in general is not a recipe for success.

We want to stress that superpositions play no fundamental role in classifying states: any state can be expressed as a linear combination of any other state. In a spin 1/2 system, for example, spin left is a linear combination of spin up and spin down with $|c_1|^2 = |c_2|^2 = \frac{1}{2}$. But spin up is also a linear combination of spin left and spin right with $|c_1|^2 = |c_2|^2 = \frac{1}{2}$. Any spin 1/2 state, in fact, can always be expressed as a linear combination of any two distinct states, not necessarily orthogonal. The reason is that any two linearly independent vectors of a two dimensional space will span the entire space.

We also want to stress that the linearity of superpositions is different from the linearity of statistical mixture. The first linearity acts on pure states only, uses complex coefficient, returns a pure state and is specific to quantum mechanics. The second linearity acts on all states, mixed or pure, uses real coefficient, returns mixed states and is present in both classical and quantum mechanics. In fact, it must be present in any physical theory, since all measurements are statistical in nature. If we pick two pure states ψ and ϕ on the Bloch sphere for a spin 1/2 system, they will be represented by two points on the surface. The set of all superposition $c_\psi|\psi\rangle + c_\phi|\phi\rangle$ is the whole surface. The set of all statistical mixtures $p|\psi\rangle\langle\psi| + (1-p)|\phi\rangle\langle\phi|$, instead, is the line segment that contains the two points. Therefore, it is not true that superpositions are "a kind of" statistical mixture.

The two linearities, however, are related and understanding their relationship is key to understand why superpositions are possible in quantum mechanics but not in classical mechanics. Mathematically, if $|\phi\rangle$ is a superposition of $|\psi_1\rangle$ and $|\psi_2\rangle$, then we can find a statistical mixture ρ of $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$ such that it can also be expressed as a mixture of $|\phi\rangle\langle\phi|$ and another state $|\hat{\phi}\rangle\langle\hat{\phi}|$. Let $|\phi\rangle = c_1|\psi_1\rangle + c_2|\psi_2\rangle$. We define $|\hat{\phi}\rangle = c_1|\psi_1\rangle - c_2|\psi_2\rangle$. We have:

$$\begin{aligned}
|\phi\rangle\langle\phi| &= c_1^* c_1|\psi_1\rangle\langle\psi_1| + c_1^* c_2|\psi_1\rangle\langle\psi_2| + c_2^* c_1|\psi_2\rangle\langle\psi_1| + c_2^* c_2|\psi_2\rangle\langle\psi_2| \\
|\hat{\phi}\rangle\langle\hat{\phi}| &= c_1^* c_1|\psi_1\rangle\langle\psi_1| - c_1^* c_2|\psi_1\rangle\langle\psi_2| - c_2^* c_1|\psi_2\rangle\langle\psi_1| + c_2^* c_2|\psi_2\rangle\langle\psi_2| \\
|\phi\rangle\langle\phi| + |\hat{\phi}\rangle\langle\hat{\phi}| &= 2|c_1|^2|\psi_1\rangle\langle\psi_1| + 2|c_2|^2|\psi_2\rangle\langle\psi_2|
\end{aligned} \tag{3.8}$$

$$\begin{aligned}
p_\phi &= \frac{\langle\phi|\phi\rangle}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle} \qquad p_{\hat{\phi}} = \frac{\langle\hat{\phi}|\hat{\phi}\rangle}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle} \\
p_1 &= \frac{2|c_1|^2\langle\psi_1|\psi_1\rangle|}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle} \quad p_2 = \frac{2|c_2|^2\langle\psi_2|\psi_2\rangle|}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle}
\end{aligned} \tag{3.9}$$

$$\begin{aligned}
\rho = p_\phi\rho_\phi + p_{\hat{\phi}}\rho_{\hat{\phi}} &= \frac{\langle\phi|\phi\rangle}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle}\frac{|\phi\rangle\langle\phi|}{\langle\phi|\phi\rangle} + \frac{\langle\hat{\phi}|\hat{\phi}\rangle}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle}\frac{|\hat{\phi}\rangle\langle\hat{\phi}|}{\langle\hat{\phi}|\hat{\phi}\rangle} \\
&= \frac{1}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle}\left(|\phi\rangle\langle\phi| + |\hat{\phi}\rangle\langle\hat{\phi}|\right) \\
&= \frac{1}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle}\left(2|c_1|^2|\psi_1\rangle\langle\psi_1| + 2|c_2|^2|\psi_2\rangle\langle\psi_2|\right) \\
&= \frac{2|c_1|^2\langle\psi_1|\psi_1\rangle|}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle}\frac{|\psi_1\rangle\langle\psi_1|}{\langle\psi_1|\psi_1\rangle} + \frac{2|c_2|^2\langle\psi_2|\psi_2\rangle|}{\langle\phi|\phi\rangle + \langle\hat{\phi}|\hat{\phi}\rangle}\frac{|\psi_2\rangle\langle\psi_2|}{\langle\psi_2|\psi_2\rangle} \\
&= p_1\rho_{\psi_1} + p_2\rho_{\psi_2}
\end{aligned} \tag{3.10}$$

Note that $p_\phi + p_{\hat{\phi}} = 1$, which means ρ is a statistical mixture with the given probability. This also means that $p_1 + p_2 = 1$ as well. We therefore have a mixed state that can be expressed either as a mixture of ψ_1 and ψ_2 or of ϕ and $\hat{\phi}$.

The converse is also true. Suppose that $\rho = p_\phi\rho_\phi + p_{\hat{\phi}}\rho_{\hat{\phi}} = p_1\rho_{\psi_1} + p_2\rho_{\psi_2}$. Since ρ is the mixture of two pure states, it is supported by a two dimensional subspace. Since ρ is a mixture of ρ_{ψ_1} and ρ_{ψ_2}, the subspace is the span of ψ_1 and ψ_2. Since ϕ must be in that subspace as well, it must be a superposition of ψ_1 and ψ_2. This means that ϕ is a superposition of ψ_1 and ψ_2 if and only if there is mixed state ρ that can be expressed as a mixture of both ψ_1 and ψ_2 as well as a mixture of ϕ and some other state. This lead to the following:

Insight 3.11. *The existence of superpositions of pure states in quantum mechanics is equivalent to the existence of multiple decomposition of statistical mixtures in terms of pure states.*

This links the abstract operation of quantum superposition to the more concrete physical property of statistical mixing.

This explains why there are no superpositions in classical mechanics. Every classical ensemble has a unique decomposition in terms of pure states, where the pure states are limits of distributions all concentrated at a point in phase space. Since classical mechanics does not allow multiple decompositions, it does not allow superpositions either.

Note that the previous insight does not tell us that if a theory allows multiple decompositions then it must be equivalent to quantum mechanics. For example, if the state space of statistical mixture is a disc instead of a ball, we would still have multiple decompositions but no complex superpositions. It does tell us, however, that what superpositions are allowed is fixed by what mixtures are allowed, which means that the pure states are rays of a complex Hilbert space if and only if the space of statistical mixtures is isomorphic to the quantum one.

The above insight also tells us why quantum mechanics is a linear theory. If we have a physical process $\mathcal{P}$ that takes as input a statistical mixture, we expect that the output of the statistical mixture is the statistical mixture of the individual outputs. That is, $\mathcal{P}(p\rho_1 + (1-p)\rho_2) = p\mathcal{P}(\rho_1) + (1-p)\mathcal{P}(\rho_2)$. But if the process is linear with respect to statistical mixing, and the linearity of statistical mixing can be converted to the linearity of superpositions, then the process must be also linear with respect to superpositions. Similarly, if we have an observable O and a statistical mixture, we expect that the expectation of the statistical mixture is the average of the expectations on the statistical components. That is $E[p\rho_1 + (1-p)\rho_2] = pE[\rho_1] + (1-p)E[\rho_2]$. In quantum mechanics, this means that $\operatorname{tr}(O(p\rho_1 + (1-p)\rho_2)) = p\operatorname{tr}(O\rho_1) + (1-p)\operatorname{tr}(O\rho_2)$. That is, the linearity of the observables is exactly the linearity of expectation values.

The gap between the two linearities

Note, however, that while the linearity of statistical mixture is a physically necessary requirement, the linearity of superposition is a choice dictated by mathematical convenience. As we saw, it is the Born rule that captures the physics, not the inner product. Therefore, a non-linear transformation that preserves the inner product is still physically valid. For example, consider the Bloch ball. Taken two orthogonal states ψ and ϕ, every other state can be expressed as a linear combination $c_\psi|\psi\rangle + c_\phi|\phi\rangle$. Now consider the map $m : \mathcal{H} \to \mathcal{H}$ defined by

$$m(c_\psi|\psi\rangle + c_\phi|\phi\rangle) \mapsto e^{\imath\theta \frac{|c_\phi|}{\sqrt{|c_\psi|^2+|c_\phi|^2}}}(c_\psi|\psi\rangle + c_\phi|\phi\rangle), \tag{3.12}$$

where $\theta \in (0, 2\pi)$ is an arbitrary constant. The map is introducing a phase shift that is proportional to the cosine of the angle along the vertical direction. Since it is introducing only a total phase shift, it will map a vector to a physically equivalent vector. The physics will not change, and we do not expect the Born rule to change. However, since the shift is not linear, we will not expect the inner product to me conserved since only linear maps can preserve the inner product.

Let's verify all of this in the math. First, the map is colinear, meaning that it maps a ray to another ray. That is:

$$\begin{aligned} m(k(c_\psi|\psi\rangle + c_\phi|\phi\rangle)) &= m(kc_\psi|\psi\rangle + kc_\phi|\phi\rangle) = e^{\imath\theta\frac{|kc_\phi|}{\sqrt{|kc_\psi|^2+|kc_\phi|^2}}}(kc_\psi|\psi\rangle + kc_\phi|\phi\rangle) \\ &= ke^{\imath\theta\frac{|k||c_\phi|}{|k|\sqrt{|c_\psi|^2+|c_\phi|^2}}}(c_\psi|\psi\rangle + c_\phi|\phi\rangle) = ke^{\imath\theta\frac{|c_\phi|}{\sqrt{|c_\psi|^2+|c_\phi|^2}}}\left(c_\psi|\psi\rangle + c_\phi|\phi\rangle\right) \\ &= k\, m(c_\psi|\psi\rangle + c_\phi|\phi\rangle) \end{aligned} \tag{3.13}$$

Moreover, we can see that each ray is mapped to the same ray. We can verify that the map in not linear. In fact:

$$\begin{aligned} m(|\psi\rangle) &= |\psi\rangle \\ m(|\phi\rangle) &= e^{\imath\theta}|\phi\rangle \\ m(|\psi\rangle + |\phi\rangle) &= e^{\frac{\imath\theta}{\sqrt{2}}}(|\psi\rangle + |\phi\rangle) \neq |\psi\rangle + e^{\imath\theta}|\phi\rangle. \end{aligned} \tag{3.14}$$

The inner product becomes:

$$\langle m(v)|m(w)\rangle = e^{-\imath\theta\frac{|v_\phi|}{\sqrt{|v_\psi|^2+|v_\phi|^2}}} e^{\imath\theta\frac{|w_\phi|}{\sqrt{|w_\psi|^2+|w_\phi|^2}}}\langle v|w\rangle, \tag{3.15}$$

which is manifestly not conserved. However, for the Born rule we have:

$$\begin{aligned} \langle m(v)|m(v)\rangle &= e^{-\imath\theta\frac{|v_\phi|}{\sqrt{|v_\psi|^2+|v_\phi|^2}}} e^{\imath\theta\frac{|v_\phi|}{\sqrt{|v\psi|^2+|v_\phi|^2}}}\langle v|v\rangle = \langle v|v\rangle \\ p(m(v)|m(w)) &= \frac{\langle m(v)|m(w)\rangle\langle m(w)|m(v)\rangle}{\langle m(v)|m(v)\rangle\langle m(w)|m(w)\rangle} \\ &= \frac{e^{-\imath\theta\frac{|v_\phi|}{\sqrt{|v_\psi|^2+|v_\phi|^2}}} e^{\imath\theta\frac{|w_\phi|}{\sqrt{|w_\psi|^2+|w_\phi|^2}}}\langle v|w\rangle e^{-\imath\theta\frac{|w_\phi|}{\sqrt{|w_\psi|^2+|w_\phi|^2}}} e^{\imath\theta\frac{|v_\phi|}{\sqrt{|v_\psi|^2+|v_\phi|^2}}}\langle w|v\rangle}{\langle v|v\rangle\langle w|w\rangle} \\ &= \frac{\langle v|w\rangle\langle w|v\rangle}{\langle v|v\rangle\langle w|w\rangle} = p(v|w). \end{aligned} \tag{3.16}$$

As expected, the Born rule is conserved while the inner product is not.

This tells us that, technically, the linearity of the underlying vector space is not a physical requirement and, in principle, we could represent pure states with a space where superposition as "slightly" non-linear. That is, the fact that the space of pure states is a complex projective space is a physical requirement if we want to represent a set of states that provides the statistical mixtures given by quantum mechanics. However, using the underlying complex inner product space is not a physical requirement: it is mathematical convenience.

Insight 3.17. *The linearity of statistical mixing is a physical requirement. The linearity of the Hilbert space is a mathematical convenience.*

Basis and gauge choice

Once we decide to use the vector space, the representation of states in terms of vector is also defined up to an arbitrary choice. In a spin 1/2 system, for example, we typically define

$|x^+\rangle = \frac{1}{\sqrt{2}}|z^+\rangle + \frac{1}{\sqrt{2}}|z^-\rangle$. But note the state corresponding to spin down is the whole ray, so the above expression hides that fact that $|z^-\rangle$ hides a choice of arbitrary phase. We could, in fact, redefine $|z^-\rangle$ with a quarter of a turn phase shift, and have $|x^+\rangle = \frac{1}{\sqrt{2}}|z^+\rangle + \imath\frac{1}{\sqrt{2}}|z^-\rangle$. In other words, the choice of representation of the physics in terms of the Hilbert space is up to an arbitrary choice of phase for elements of the basis.

For continuous observable, this choice is even more important, as it is not a matter of basis, but of operators. Suppose we have a wave function $\psi(x)$. The corresponding position and momentum operators will be $X = x$ and $P = -\imath\hbar\partial_x$. Now choose an arbitrary continuous function $\theta(x)$. This corresponds to an arbitrary choice of phase at each point. If we set $\phi(x) = e^{\imath\theta(x)}\psi(x)$, the probability distribution over position will not change as the norm at each point remains the same. If set $P_\phi = -\imath\hbar\partial_x - \hbar\partial_x\theta(x)$, we have

$$\{X, P_\phi\} = \{X, P - \hbar\partial_x\theta(x)\} = \{X, P\} - \{X, \hbar\partial_x\theta(x)\} = \{X, P\} \tag{3.18}$$

since X will commute with any function of x. This means that X and P_θ have the same commutation relationships than X and P. Also note that

$$\begin{aligned}
\langle\phi|P_\phi|\phi\rangle &= \int_X e^{-\imath\theta}\psi(x)(-\imath\hbar\partial_x - \hbar\partial_x\theta(x))(e^{\imath\theta}\psi(x))dx \\
&= \int_X e^{-\imath\theta}\psi(x)(-\imath\hbar e^{\imath\theta}\imath\partial_x\theta(x)\psi(x) - e^{\imath\theta}\imath\hbar\partial_x\psi(x) - \hbar\partial_x\theta(x)e^{\imath\theta}\psi(x))dx \\
&= \int_X e^{-\imath\theta}\psi(x)(-e^{\imath\theta}\imath\hbar\partial_x\psi(x))dx \\
&= \int_X \psi(x)(-\imath\hbar\partial_x)\psi(x)dx = \langle\psi|P_\psi|\psi\rangle.
\end{aligned} \tag{3.19}$$

In other words, we can redefine both the phase of the wave function and the momentum operator so that the physics remains unchanged. This corresponds to the change of gauge we had in classical particle mechanics, where we could redefine momentum $p \to p + \partial f(x)$ while leaving the Poisson brackets, and therefore the space-time trajectories, unchanged.

Insight 3.20. *The choice of a basis in the Hilbert space is a choice of a maximal set of orthogonal states plus a choice of gauge (i.e. a phase for each element of the basis).*

To sum up, fixing the Hilbert space and its physical representation is equivalent to fixing the space of ensembles, including how they mix, requiring a linear representation in terms of the Hilbert space and choosing an arbitrary phase for each element of the spectrum of a complete set of observables.

Probability from expectations

Another interesting equivalence is the one between probability, subspaces and expectations. This equivalence exists also in classical probability, therefore we should understand it there fit. The standard way to define a probability space is by giving three objects. First, a sample space Ω that represents all the possible cases. In classical mechanics, it's phase space, where each point represent all the values of position and momentum. Second, an algebra of events Σ_Ω which represents all possible conditions we could test for. In classical mechanics, this is the collection of Borel sets, subsets of phase space that can be generated from the open regions using complements and countable unions. Third, a probability measure $p : \Sigma_\Omega \to [0, 1]$ that

associated a probability to each event. From this one can define random variables $X : \Omega \to \mathbb{R}$ and their expectation. That is, we start with a notion of probability and define a notion of observables on top.

We could proceed the other way. The key insight is that for every event $A \in \Sigma_\Omega$ we can define the indicator function $\mathbf{1}_A : \Omega \to \{0, 1\}$ that returns one if the given point is in A or zero otherwise. This is a random variable and its expectation is exactly the probability of the event A. That is, $E[\mathbf{1}_A] = p(A)$. Suppose, therefore, that we are given Ω, Σ_Ω but not the probability measure. However, we are given the expectation operator $E : M(\Omega, \mathbb{R})$ that, given a function $f : \Omega \to \mathbb{R}$[2], $E[f]$ returns it's expectation. We can then define the probability measure as $p(A) = E[\mathbf{1}_A]$.

A similar perspective can be achieved in quantum mechanics, and it is the basis of some other axiomatic approaches and quantum reconstruction approaches, like quantum logic. As we are using the same notation for indicator functions and projectors, it should not be surprising that projectors play the role of indicator function in quantum mechanics. A projector P is an idempotent linear operator, meaning that $P^2 = P$. This also means that $E[P^2] = E[P]$. This is true for an indicator function I as well: since the only values I takes are zeros and one, $I^2 = I$, and $E[I^2] = E[I]$. On the other hand, the spectrum of a projector (i.e. the possible eigenvalues) consists only of zero and one, just like an indicator function. Therefore, if we can reverse the definitions for probability in quantum mechanics as well.

The standard construction starts with a set of pure states $\mathbf{P}(\mathcal{H})$, defines the Born rule $p(\phi|\psi)$ in terms of the inner product, and then the expectation based on the Born rule. We can start with a set of pure states $\mathbf{P}(\mathcal{H})$, where $\mathcal{H}$ is taken as a vector space forgetting its inner product, the space of all possible observables $O(\mathcal{H})$ and the expectation operator $E : O(\mathcal{H}) \to \mathbb{R}$. For every pure state $\phi \in \mathbf{P}(\mathcal{H})$ we can find the corresponding projector $\mathbf{1}_\phi$ and define $p(\phi) = E[\mathbf{1}_\phi]$. This allows to reconstruct the mixed state, much like we reconstructed the probability measure in classical mechanics. To reconstruct the Born rule, for each pure $\psi \in \mathbf{P}(\mathcal{H})$ we need an E_ψ that gives us the expectations of all the variables if we prepared ψ. Then we can set $p(\phi|\psi) = E_\psi[\mathbf{1}_\phi]$.

Therefore condition

Expectations are defined for all observables and for all ensembles (BR-EXP)

is an alternative characterization of the Born rule.

Born rule and angles in projective space

As we saw, the Born rule is essentially a relationship between rays, points in the projective space. We want to understand what exactly is it describing. Since it is a relationship between two pure states, let us take two rays $\psi, \phi \in P(\mathcal{H})$. Regardless of the dimensionality of $\mathcal{H}$, ψ and ϕ will span a two-dimensional subspace, a Bloch sphere. In fact, considering the circle that contains ψ, ϕ and its opposite $\phi_\perp$ will suffice.

Since ϕ and $\phi_\perp$ are opposite in the Bloch sphere, and therefore orthogonal, we can write

[2]The function will need to be measureable in the mathematical sense: the inverse image of a Borel set is a Borel set.

$|\psi\rangle = \rho|\phi\rangle + \sqrt{1-\rho^2}e^{i\theta}|\phi_\perp\rangle$. We have

$$p(\psi|\phi) = \frac{\langle\phi|\psi\rangle\langle\psi|\phi\rangle}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle} = \rho^2 = \cos^2\theta_{\psi\phi}$$
$$\theta_{\psi\phi} = \arccos\sqrt{\frac{\langle\phi|\psi\rangle\langle\psi|\phi\rangle}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle}}. \tag{3.21}$$

In other words, the Born rule defines the geometry of the projective space by defining the angles between points. Note that, since we are on a unit sphere, the angle can be understood also as a distance along the maximal circle that connects the two. The maximal circles are the geodesics of the sphere and therefore the angle is the distance between the points. Therefore $\theta_{\psi\phi}$ is literally the distance function that define the geometry of the space.[3] We have that

The angles between pure states are well defined. (BR-ANG)

is an alternative characterization of the Born rule.

Entropy and the Born rule

TODO: preamble.

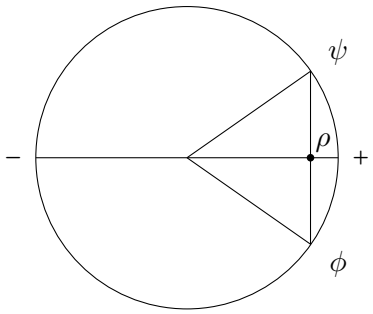


Figure 3.1: Geometric diagonalization of an equal mixture of two states.

To calculate the entropy of a mixture of two pure states, we can use the understanding that we built in terms of ensembles and their multiple decomposition. Let ψ and ϕ be two pure states. Their equal mixture $\rho = \frac{1}{2}\rho_\psi + \frac{1}{2}\rho_\phi$ will be the midpoint between them. To calculate the entropy, we need to diagonalize ρ and calculate the Shannon entropy from the eigenvalues. To diagonalize ρ we need to express it as a mixture of two orthogonal + and −. But + and − will be orthogonal only if they are opposite points on the Bloch sphere. Therefore, ρ is diagonalized by + and − if and only if it lies on the axis identified by + and − as we see in figure 3.1. Geometrical, we will have $\rho = \frac{\overline{-\rho}}{\overline{-+}}\rho_+ + \frac{\overline{\rho+}}{\overline{-+}}\rho_-$. Using simple trigonometry and recalling the relationship between the Born rule and angles, we have:

$$\cos\theta_{\psi+} = \cos\frac{\theta_{\psi\phi}}{2} = \sqrt{p(\phi|\psi)}$$
$$\rho = \frac{\overline{-\rho}}{\overline{-+}}\rho_+ + \frac{\overline{\rho+}}{\overline{-+}}\rho_- = \frac{1+\cos\theta_{\psi+}}{2}\rho_+ + \frac{1-\cos\theta_{\psi+}}{2}\rho_- \tag{3.22}$$
$$= \frac{1+\sqrt{p(\phi|\psi)}}{2}\rho_+ + \frac{1-\sqrt{p(\phi|\psi)}}{2}\rho_-.$$

[3] The angle is the distance as calculated using the Fubini-Study metric.

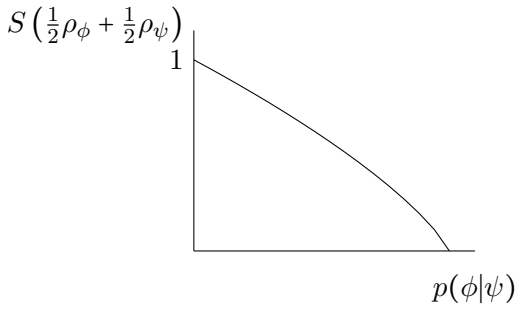


Figure 3.2: Entropy of a mixture of two pure states as a function of their Born rule.

The entropy is therefore

$$S(\rho) = -\operatorname{tr}(\rho \log \rho) = -\frac{1+\sqrt{p(\phi|\psi)}}{2}\log\frac{1+\sqrt{p(\phi|\psi)}}{2} - \frac{1-\sqrt{p(\phi|\psi)}}{2}\log\frac{1-\sqrt{p(\phi|\psi)}}{2}. \quad (3.23)$$

The relationship is plotted in figure 3.2 and, as one can see, it is strictly concave. The relationship is, therefore, invertible: if we are given the entropy of all pairs of state we would be able to recover the Born rule. In other words

The entropy is well defined for all ensembles (BR-ENT)

is an alternative characterization of the Born rule.

States are rays ¡-¿ states are pure ensembles.
Probability as expectations of indicator functions
Projections as indicator functions
Orthogonality as mutual exclusivity
Inner product as overlap between pure ensembles

3.3 Schroedinger equation and unitary evolution

We start with the Schroedinger equation

The evolution follows the equation $\imath\hbar\frac{d}{dt}\psi(t) = H\psi(t)$ where H is a self-adjoint operator (DR-SCEQ)

Time evolution is unitarity

The Schroedinger equation describes time evolution as a relationship between the change of the state and the Hamiltonian of the system. Here we want to characterize time evolution as a map U_{dt} that takes the state at time t and returns the state at time $t + dt$. We have

$$\begin{aligned}\psi(t+dt) &= \psi(t) + d\psi(t) = \psi(t) + \frac{d}{dt}\psi(t)dt = \psi(t) + \frac{Hdt}{\imath\hbar}\psi(t)\\ &= \left(1 + \frac{Hdt}{\imath\hbar}\right)\psi(t) = U_{dt}\psi(t).\end{aligned} \quad (3.24)$$

Since H is a self-adjoint operator, we have

$$\begin{aligned} U_{dt}^\dagger U_{dt} &= \left(1+\frac{H dt}{\imath\hbar}\right)^\dagger \left(1+\frac{H dt}{\imath\hbar}\right) = \left(1+\frac{H^\dagger dt}{(-\imath)\hbar}\right)\left(1+\frac{H dt}{\imath\hbar}\right) = \left(1-\frac{H dt}{\imath\hbar}\right)\left(1+\frac{H dt}{\imath\hbar}\right) \\ &= 1+\left(\frac{H dt}{\imath\hbar}\right)^2 = 1+O(dt^2). \end{aligned} \tag{3.25}$$

This means that the infinitesimal time evolution operator U_{dt} is unitary.

We can proceed in the opposite way. Given an infinitesimal unitary time evolution operator, we can recover the change of the state in time.

$$\begin{aligned} \frac{d\psi(t)}{dt} &= \frac{\psi(t+dt)-\psi(t)}{dt} = \frac{U_{dt}\psi(t)-\psi(t)}{dt} = \frac{U_{dt}-1}{dt}\psi(t) = A\psi(t) \\ U_{dt} &= 1+A dt \end{aligned} \tag{3.26}$$

From the unitarity we have

$$\begin{aligned} U_{dt}^\dagger U_{dt} &= (1+A dt)^\dagger (1+A dt) = \left(1+A^\dagger dt\right)(1+A dt) = 1+\left(A+A^\dagger\right) dt + A^\dagger A dt^2 \\ &= 1+\left(A+A^\dagger\right) dt + O(dt^2) = 1 \\ A &= -A^\dagger \end{aligned} \tag{3.27}$$

Therefore A is skew-adjoint, and we can set $A = \frac{H}{\imath\hbar}$ where H is a self-adjoint operator. Therefore $\imath\hbar\frac{d}{dt}\psi(t) = \imath\hbar A\psi(t) = H\psi(t)$. The evolution follows the Schroedinger equation. To sum it up

Time evolution is unitary: $U_{dt}^\dagger U_{dt} = U_{dt}U_{dt}^\dagger = 1$ (DR-UNIT)

is equivalent to DR-SCEQ.

TODO: make sure to discuss the invertibility condition!

Time evolution is preservation of the inner product

Note that if the evolution is unitary, then the inner product is conserved. In fact

$$\langle U_{dt}\phi | U_{dt}\psi\rangle = \langle \phi | U_{dt}^\dagger U_{dt} | \psi\rangle = \langle \phi|\psi\rangle. \tag{3.28}$$

The argument works in reverse as well: an infinitesimal time evolution operator that conserves the inner product is unitary. Therefore

Time evolution preserves the inner product : $\langle U_{dt}\phi | U_{dt}\psi\rangle = \langle\phi|\psi\rangle$ (DR-INN)

is equivalent to DR-UNIT.

Time evolution is preservation of the norm

In particular, unitary evolution will preserve the norm of vectors, since $|\psi|^2 = \langle\psi|\psi\rangle$. Note that the inner product can be expressed, through the polarization identity, in terms of the norm.

$$\langle\phi|\psi\rangle = \frac{1}{4}\left(|\phi+\psi|^2 - |\phi-\psi|^2 - \imath|\phi+\imath\psi|^2 + \imath|\phi-\imath\psi|^2\right) \tag{3.29}$$

If U_{dt} is a linear transformation that preserves the norm, that is $|U_{dt}\psi|^2 = |\psi|^2$, then, we have

$$\begin{aligned}\langle U_{dt}\phi|U_{dt}\psi\rangle &= \frac{1}{4}\left(|U_{dt}\phi + U_{dt}\psi|^2 - |U_{dt}\phi - U_{dt}\psi|^2 - \imath|U_{dt}\phi + \imath U_{dt}\psi|^2 + \imath|U_{dt}\phi - \imath U_{dt}\psi|^2\right)\\ &= \frac{1}{4}\left(|U_{dt}(\phi+\psi)|^2 - |U_{dt}(\phi-\psi)|^2 - \imath|U_{dt}(\phi+\imath\psi)|^2 + \imath|U_{dt}(\phi-\imath\psi)|^2\right)\\ &= \frac{1}{4}\left(|\phi+\psi|^2 - |\phi-\psi|^2 - \imath|\phi+\imath\psi|^2 + \imath|\phi-\imath\psi|^2\right) = \langle\phi|\psi\rangle\end{aligned} \tag{3.30}$$

Therefore condition

> Time evolution is linear and preserves the norm: $\langle\psi(t)|\psi(t)\rangle = \langle\psi(t+dt)|\psi(t+dt)\rangle$ (DR-NORM)

is equivalent to condition DR-INN.

Time evolution is preservation of the Born rule

We now look at the square of the inner product between two states infinitesimally close in time. We have

$$\begin{aligned}|\langle\psi(t)|\psi(t+dt)\rangle|^2 &= \langle\psi(t)|U_{dt}\psi(t)\rangle\langle U_{dt}\psi(t)|\psi(t)\rangle\\ &= \langle\psi(t)|1 + \frac{Hdt}{\imath\hbar}|\psi(t)\rangle\langle\psi(t)|\left(1+\frac{Hdt}{\imath\hbar}\right)^\dagger|\psi(t)\rangle\\ &= \langle\psi(t)|1 + \frac{Hdt}{\imath\hbar}|\psi(t)\rangle\langle\psi(t)|1 - \frac{Hdt}{\imath\hbar}|\psi(t)\rangle\\ &= \langle\psi(t)|\psi(t)\rangle\langle\psi(t)|\psi(t)\rangle + \langle\psi(t)|\psi(t)\rangle\langle\psi(t)|-\frac{Hdt}{\imath\hbar}|\psi(t)\rangle\\ &+ \langle\psi(t)|\frac{Hdt}{\imath\hbar}|\psi(t)\rangle\langle\psi(t)|\psi(t)\rangle + O(dt^2)\\ &= |\langle\psi(t)|\psi(t)\rangle|^2 + O(dt^2).\end{aligned} \tag{3.31}$$

In particular, if the vector is normalized, we have

$$|\langle\psi(t)|\psi(t+dt)\rangle|^2 = 1 + O(dt^2). \tag{3.32}$$

That is, the square of the inner product between two infinitesimally close states is one.

For the converse, let's assume U_{dt} is such that $|\langle\psi(t)|\psi(t+dt)\rangle|^2 = |\langle\psi(t)|\psi(t)\rangle|^2$. As we saw before, since U_{dt} is infinitesimal, we can write $U_{dt} = 1 + Adt$. We have:

$$\begin{aligned}|\langle\psi(t)|\psi(t)\rangle|^2 &= |\langle\psi(t)|\psi(t+dt)\rangle|^2 = \langle\psi(t)|U_{dt}\psi(t)\rangle\langle U_{dt}\psi(t)|\psi(t)\rangle\\ &= \langle\psi(t)|U_{dt}|\psi(t)\rangle\langle\psi(t)|U_{dt}^\dagger|\psi(t)\rangle\\ &= \langle\psi(t)|1+Adt|\psi(t)\rangle\langle\psi(t)|1+A^\dagger dt|\psi(t)\rangle\\ &= |\langle\psi(t)|\psi(t)\rangle|^2 + \langle\psi(t)|\psi(t)\rangle\langle\psi(t)|(A+A^\dagger)dt|\psi(t)\rangle + O(dt^2).\end{aligned} \tag{3.33}$$

This means that $A = A^\dagger$ and therefore $U_{dt} = 1 + Adt$ is unitary. Therefore condition

> The square of the inner product between to states infinitesimally close in time is one: $|\langle\psi(t)|\psi(t+dt)\rangle|^2 = 1$ (DR-UBOR)

is yet another equivalent condition to unitary evolution.

States and states change are perpendicular on the complex plane

The above two conditions make sense if we understand what happens geometrically. A unitary evolution is effectively a rotation in the Hilbert space, that is why the norm is conserved. An infinitesimal unitary evolution, then, is an infinitesimal rotation so the change is tangent to the circle, and therefore perpendicular with respect to the original vector. This is why the projection of the new vector onto the old one is equal to the norm.

However, perpendicular does not mean orthogonal in this case. In the complex plane, a multiplication by $\imath$ gives us a perpendicular vector that is not orthogonal with respect to the inner product. To see this

$$\begin{aligned}\langle\psi(t+dt)|\psi(t+dt)\rangle &= \langle\psi(t)+d\psi|\psi(t)+d\psi\rangle \\ &= \langle\psi(t)|\psi(t)\rangle + \langle\psi(t)|d\psi\rangle + \langle d\psi|\psi(t)\rangle + \langle d\psi|d\psi\rangle \\ &= \langle\psi(t)|\psi(t)\rangle + \langle\psi(t)|Adt|\psi(t)\rangle + \langle\psi(t)|A^\dagger dt|\psi(t)\rangle + O(dt^2) \\ \langle\psi(t+dt)|\psi(t+dt)\rangle - \langle\psi(t)|\psi(t)\rangle &= \langle\psi(t)|(A+A^\dagger)dt|\psi(t)\rangle + O(dt^2)\end{aligned} \tag{3.34}$$

Note how $A+A^\dagger$ gives us the change in the norm of ψ during the evolution. The norm does not change if and and only if A is skew-adjoint. Since $\langle\psi(t)|d\psi(t)\rangle = \langle\psi(t)|A|\psi(t)\rangle \neq 0$, the change in not orthogonal in the Hilber space. However, the quantity is imaginary so the change is perpendicular in the complex plane of every dimension. If we consider the triangle formed by $\psi(t+dt)$, $\psi(t)$ and $d\psi$, the triangle is perpendicular if and only if A is skew-adjoint. Therefore the condition

> The change is perpendicular to the original state: $\langle\psi(t)|d\psi(dt)\rangle$ is imaginary (DR-PERP)

is yet another equivalent condition to DR-SCEQ.

Time evolution is preservation of bases orthonormality

Another way to characterize unitary evolution is through what happens to an orthonormal basis $|e_i\rangle$. Given that a unitary evolution preserves the inner product, we have $\langle U_{dt}e_i|U_{dt}e_j\rangle = \langle e_i|e_j\rangle = \delta_{ij}$. Therefore the unitary evolution maps an orthonormal basis to another orthonormal basis. The converse is also true, if U_{dt} is linear and maps an orthonormal basis to another orthonormal basis, then the inner product is preserves. To see this, we can simply expand any vector in terms of the basis vector. That is

$$\begin{aligned}\langle\psi(t)|\phi(t)\rangle &= \langle c_i e_i|d_j e_j\rangle = c_i^* d_j\langle e_i|e_j\rangle = c_i^* d_j\langle U_{dt}e_i|U_{dt}e_j\rangle \\ &= \langle U_{dt}c_i e_i|U_{dt}d_i e_j\rangle = \langle U_{dt}\psi(t)|U_{dt}\phi(t)\rangle = \langle\psi(t+dt)|\phi(t+dt)\rangle\end{aligned} \tag{3.35}$$

Therefore condition

> Time evolution is linear and maps orthonormal basis to orthonormal basis: $\langle U_{dt}e_i|U_{dt}e_j\rangle = \langle e_i|e_j\rangle = \delta_{ij}$ (DR-OBAS)

The above condition can be relaxed to just preserving pairs of orthonormal vectors. Take any two vectors ψ and ϕ, not necessarily orthogonal. We can write $\phi_\perp = \psi - \frac{\langle\phi|\psi\rangle}{\langle\phi|\phi\rangle}\phi$ and $\psi_\perp = \phi - \frac{\langle\psi|\phi\rangle}{\langle\psi|\psi\rangle}\psi$. Note that $\langle\phi|\phi_\perp\rangle = 0$ and $\langle\psi|\psi_\perp\rangle = 0$. We have

$$\begin{aligned}\langle U_{dt}\phi|U_{dt}\psi\rangle &= \langle U_{dt}\phi|U_{dt}(\phi_\perp + \frac{\langle\phi|\psi\rangle}{\langle\phi|\phi\rangle}\phi)\rangle = \langle U_{dt}\phi|U_{dt}\phi_\perp\rangle + \frac{\langle\phi|\psi\rangle}{\langle\phi|\phi\rangle}\langle U_{dt}\phi|U_{dt}\phi\rangle \\ &= 0 + \frac{\langle\phi|\psi\rangle}{\langle\phi|\phi\rangle}\langle\phi|\phi\rangle = \langle\phi|\psi\rangle,\end{aligned} \tag{3.36}$$

which means the inner product is preserved.

Physical conditions

In classical mechanics we saw that Hamiltonian evolution was equivalent to determinism and reversibility. The same applies to quantum mechanics, as we will see by looking at different but equivalent conditions.

Determinism and Reversibility

Suppose we start with a pure state $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$, meaning that the ψ is prepared with 100% probability. If U_{dt} is unitary, $\rho(t+\Delta t) = U_{dt}|\psi(t)\rangle\langle\psi(t)|U_{dt}^\dagger = |U_{dt}\psi(t)\rangle\langle U_{dt}\psi(t)| = |\psi(t+\Delta t)\rangle\langle\psi(t+\Delta t)|$, which means the final state is also a pure state. There is a one-to-one correspondence between initial and finial state, and therefore the evolution is deterministic and reversible. Conversely, if the evolution is deterministic and reversible, if we start with a pure state $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$, then the final state will have to be $\rho(t+\Delta t) = |\psi(t+\Delta t)\rangle\langle\psi(t+\Delta t)|$. Given that both $\rho(t)$ and $\rho(t+\Delta t)$ are trace one operators, the norm of both $\psi(t)$ and $\psi(t+\Delta t)$ must be unitary. That is, a deterministic and reversible evolution must preserve the norm. Which means that

The evolution is deterministic and reversible (DR-EV)

is equivalent to DR-SCEQ.

Preservation of probability distributions

The same argument can be developed on probability distributions, similarly to what we have seen in classical mechanics. If we start with a probability distribution, the final probability distribution must be the same as all the probability for one case must be mapped and only mapped to a single other case. That is, suppose that we start with a mixed state $\rho(t) = p_i|e_i(t)\rangle\langle e_i(t)|$, meaning that it can be understood as a classical mixture of a set of orthogonal states. If we have a deterministic and reversibile evolution, the final state must be $\rho(t+\Delta t) = p_i|U_{dt}e_i(t)\rangle\langle U_{\Delta t}e_i(t)|$. This is the case if and only if time evolution maps an orthonormal basis to an orthonormal basis. This means the evolution is unitary and

The evolution preserves probability distributions (DR-EV)

is another equivalent condition.

Preservation of information entropy

When looking at classical mechanics, we saw that determinism and reversibility could be expressed as conservation of information entropy. This is true for quantum mechanics as well. Let $\rho(t) = p_i|e_i(t)\rangle\langle e_i(t)|$. If U_{dt} is unitary, we have $\rho(t+\Delta t) = p_i|U_{dt}e_i(t)\rangle\langle U_{\Delta t}e_i(t)|$. Since the transformed orthonormal basis is still an orthonormal basis, the entropy in both cases is given by $-\sum_i p_i \log p_i$, which means it is conserved.

Conversely, if an evolution preserves entropy then pure states must be mapped to pure states because all pure states and only pure states have zero entropy. Moreover, we saw that the square of the inner product characterizes the entropy of the mixture of a pair of states, which must be conserved if entropy is to be conserved. In particular, orthogonal states must remain orthogonal since they maximize entropy increase. Therefore, an evolution that

preserves entropy is one that preserver orthonormality and therefore it is unitary. That is, condition

The evolution preserves information entropy (DR-INFO)

is equivalent to DR-SCEQ.

Quasi-static processes

Note that in quantum mechanics both information entropy and thermodynamic entropy coincide, in the sense that we do not have two different definition in quantum statistical mechanics as we have in classical statistical mechanics (i.e. logarithm of count of states and Shannon entropy). However, we saw that projectors are more fundamental as they are implied by the mere definition of a Hilbert space, and projectors can be understood as equilibration processes. In thermodynamics, reversible processes are quasi-static processes. We can show that unitary evolution is, in this sense, a quasi-static process: unitary evolution can be understood as an infinite sequence of projections that perturb the system minimally.

To give intuition, suppose that a beam of light passes through two linear polarizers, the first oriented vertically and the second horizontally. No light will pass through. Recall, in fact, that the intensity decreases by a factor of $\cos^2\varphi$ where φ is the difference in angle between the polarizers. However, if you put another polarizer in between at a 45 degree angle, then some light will have a chance to be pass through. You can put another two polarizers so that the angle between any consecutive pairs is 22.5 degrees. More light will go through. We can imagine to repeat this process, until we have a large sequence of polarizers at a small angle. In that case, $\cos^2\varphi \approx 1 - \frac{\varphi^2}{2}$. Note that, to a first order, all light will go through. Therefore, in the limit, the net effect of the polarizers is to rotate the polarization of light from vertical to horizontal. This idea generalizes.

We saw, in fact, that for a unitary evolution $\langle\psi(t+dt)|\psi(t)\rangle = 1$, that is the projection of the state at a future time step on the previous time step is one. This can be understood as making a projective measurement on an observable that is slightly different to one for which $\psi(t)$ is an eigenstate. That is, we can understand unitary evolution as an infinitesimal sequence of projections at each time step. Note that the direction of the projection depends on the initial state. This is consistent with the evolution being deterministic: if we assume that the final state is the outcome of a projective measurement, the process is deterministic if and only if the choice of projective measurement depends on the initial state.

Another way of understanding this is that determinism and reversibility can be used for both measuring and preparing states. That is, if we prepare a system in a given state, we can use unitary evolution to prepare a system in the future state. Conversely, if we can measure a system in a given state, we can use unitary evolution to infer the state of the system at a prior time. Therefore, we can understand determinism and reversibility as a series of preparations or measurements. Since measurements in quantum mechanics are projections, it makes sense that we can understand unitary evolution as a sequence of projections. Therefore

Time evolution is a quasi-static process (DR-PSEQ)

and

The evolution is an infinite sequence of reversible measurements (DR-MSEQ)

are equivalent to DR-EV.

3.4 Projection and measurements

WTS: Every state is an eigenstate of a unitary evolution, of a projection, and of an Observable.

Given a pure state $|\psi\rangle$ we can always create the operator $P_\psi = |\psi\rangle\langle\psi|$. This operator has $|\psi\rangle$ as an eigenvector with eigenvalue one, and any state $|\phi\rangle$ that is orthogonal to $|\psi\rangle$ will also be an eigenvector with eigenvalue zero. In fact:

$$\begin{aligned} P_\psi|\psi\rangle &= |\psi\rangle\langle\psi|\psi\rangle = |\psi\rangle \\ P_\psi|\phi\rangle &= |\psi\rangle\langle\psi|\phi\rangle = |\psi\rangle 0 = 0 \end{aligned} \tag{3.37}$$

Note that $P_\psi^\dagger = P_\psi$ which means that this is a Hermitian operator and since $P_\psi^\dagger P_\psi = |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = |\psi\rangle 1\langle\psi| = |\psi\rangle\langle\psi|$

$X = |x_i\rangle x_i\langle x_i|$

Projections are processes with equilibria (all fine states are equilibria) * (?) Projection is not enough: need compatility with a unitary evolution * Show that projections cannot decrease entropy * Eigenstates of projections are equlibria =¿ all quantum states are equilibria of projection - Mathematically, this is what Hilbert spaces add on top of Banach spaces * Analogy to thermodynamics (context is like different type of ensembles) * Unitary evolution is quasi-static evolution (like in thermodynamics) - Make the parallel to S-matrix calculation where we put initial state at minus infinity, and final state at plus infinity for a process that actually last "femtoseconds"

3.5 next

Observables * (?) Convex maps of mixed states are Hermitian operators * Is it useful to note that any observable is compatible to some unitary? That is, any observable is left unchanged by a unitary?

Open quantum systems * Review open quantum system (Lindblad master equation) * (?) recover CPTP maps - Linear maps : map mixed states to mixed states while preserving mixtures - Trace preserving : map trace one operators to trace one operators - Positive : mixed states have non-negative eigenvalues - Completely Positive: (?) need a characterization of completely positive in terms of only the system, without the ancilla * (?) Kraus operator, jump operatos: how are they related? If they are? * (?) What are the possible motions on a Bloch sphere? That is, what are the possible vector fields described by the Lindblad equation

Classical limit * Classical mechanics is the high entropy limit of quantum mechanics - Find classical transformations that increase entropy, show that they are all "unitaries" plus stretch of phase space - Find equivalent of phase space stretching in quantum mechanics - See that it is a CPTP map only defined in the anti-normal ordering - Show that it rescales the commutator by the factor for phase space stretching - Show that this is equivalent to the limit of $\hbar \to 0$

Negative probability in quantum mechanics * QM on phase space (Wigner functions - Hussimi - GlauberSudarshan) * (real) convex combination vs affine combination vs linear combination * =¿ QM on phase space is using affine combinations, since convex combinations are not sufficient

Quantum states as equibria * Show that for a unitary evolution, eigenstates are equilibria * Show that any quantum state is an eigenstate of some unitary * =¿ all quantum states are equilibria of unitary

(?) Recover spin 1/2 (two state systems) * Space of ensembles that is fully characterized by an average direction. - Gaussian states are fully identified by average and standard deviation - Suppose we have "guassian states" of directions with same standard deviation - -¿ the space is a ball - (?) how much can we recover? * Space of directional pure states - I have a state space for directions in space - Recycle the argument that we have to be able to put a frame invariant distribution over it - -¿ two sphere is the only symplectic sphere and therefor is the only space - (?) why are ensembles the Bloch?

3.6 Problems with infinite dimensional spaces

In the previous sections we restricted ourselves to finite dimensional spaces. In these spaces all measurements can be understood as having finitely many outcomes, and all those outcomes can be understood as quantum states. In this section we will extend the discussion to the infinite case and see the extension is problematic. Infinite dimensional Hilbert spaces, in fact, seem to hide the source of the infinity and require the existence of states that cannot be thought as being physically meaningful. The exact mathematical representation of these cases, then, is still an open problem.

There are two potential sources of infinity in physics: the infinitely large and the infinitely small. The infinitely large comes from unbounded quantities. For example, the number of particles in a gas or the distance of a particle from the origin of our reference system can be, in principle, arbitrarily large. In this case, infinity is just the range of possible values, and not a value itself. It would not make sense, for example, to say that a gas has infinitely many particles or that a particle is infinitely distant from the origin unless we are talking about the limit of a process that takes an infinite amount of time. Note that the infinitely large does not change the nature of the quantity: the number of particles is a discrete quantity and the position is a continuous quantity regardless of whether we are allowing an infinite range or not.

The infinitely small comes from the ability to refine measurements indefinitely. For example, we assume we can measure the position of a particle with arbitrary precision. While the infinite precision measurement is never realizable, the infinite precision value can be understood as the information needed to specify the finite precision outcome at all level of precision. That is, if we knew the position with infinite precision we would know all the possible finite precision intervals in which the particle can be. The infinitely small, then, changes the nature of the quantity, going from discrete to continuous. Over a finite range, a discrete quantity will have finitely many possible values while a continuous quantity will have infinitely many.

We can characterize these differences in the following way. A measurement will tell us whether a particular value is within a set of possible values. If the quantity is continuous, all measurements will always restrict the range of possible values to an infinite set. That is, any measurement of position will have a finite uncertainty, which will include infinitely many possible positions. If the quantity is discrete, at some point, we will have a measurement that identifies each possible case. That is, we can count exactly the number of particles, which restricts the measurement to only one possible case. Intuitively, the range is infinite in either cases if it cannot be always covered with finitely many measurements. That is, if we are given measurements with finite ranges, we are not going to be able to cover the infinite range with finitely many measurements.

Mathematically, this maps to properties of open sets of the topology. The topology, in fact, keeps track of the notion of closeness, and measurement resolution is about that closeness. The topology of a real line, then, is different from the topology of sets of points: the first one is topologically connected, while the second one is not. A space with a finite range is topologically compact, while one that has an infinite range is not. Establishing a perfect mapping between these concepts is not the goal of Reverse Physics, but rather of Physical Mathematics.

In classical mechanics, the mathematics characterizes and keeps track of these differences. The problem is that in quantum mechanics, the mathematical framework does not keep track of these differences. This is the problem we are going to explore in this section, a problem that is ultimately not solved. Given that the mathematical framework does not capture all the elements and only the elements that are physically meaningful, the Reverse Physics program cannot be fully completed for quantum mechanics.

Equivalence of Hilbert spaces

One feature of Hilbert spaces is that the cardinality of the base fully characterizes the space. In fact, let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces with the same cardinality and let $\{e_i\}_{i\in I} \in \mathcal{H}_1$ and $\{g_i\}_{i\in I} \in \mathcal{H}_2$ be two orthonormal basis of the respective spaces. Then we can define a map $m : \mathcal{H}_1 \to \mathcal{H}_2$ such that $m(\sum a^i e_i) = \sum a^i g_i$. The map is unitary since

$$\langle a^i e_i | a^i e_i \rangle = \langle a^i g_i | a^i g_i \rangle = \sum |a^i|^2 \tag{3.38}$$

which means the two spaces are unitarily equivalent and therefore they are the same Hilbert space.

All spaces we are interested in quantum mechanics are going to have a countable basis. If we imagine to extend to infinite the range of a discrete observable, we can see that we will have a countable set of possible outcomes and therefore a countable basis. If we take the space of wave functions, in either a finite or infinite range, we will get the space of square integrable functions, which also has a countable basis. To see that, note that the Hamiltonian for the harmonic oscillator has a discrete spectra and therefore has countably many eigenstates which will form a basis. All wave-functions, then, can be written as a superposition of eigenstates of the harmonic oscillator and therefore the space has a countable basis.

In quantum mechanics, then, if the space has an infinite basis, it will be a countable one, which leads to the following observation: all infinite dimensional spaces in quantum mechanics are equivalent. If we have an infinite dimensional space we are not going to be able to know whether we have a discrete quantity over an infinite range, a continuous quantity over a finite range or a continuous quantity over an infinite range. Mathematically, there will be no difference in the states, unlike in classical mechanics.

For example, the state space of a single DOF (i.e. $L^2(\mathbb{R})$) is going to be equivalent to the state space of n DOFs (i.e. $L^2(\mathbb{R}^n)$)). As we said before, the Hermite functions, the eigenstates of a Harmonic oscillator, provide a countable basis for a single DOF. We can imagine a Harmonic oscillator over 2 DOFs. In that case, the products of the Hermite functions across degrees of freedom provide a countable basis for the space. Through diagonalization, we can

map one set of basis to the other

$$h_0(x) \mapsto h_0(x)h_0(y) \tag{3.39}$$
$$h_1(x) \mapsto h_0(x)h_1(y) \tag{3.40}$$
$$h_2(x) \mapsto h_1(x)h_1(y) \tag{3.41}$$
$$h_3(x) \mapsto h_0(x)h_2(y) \tag{3.42}$$
$$\dots \tag{3.43}$$

This idea can be generalized to map any set of n DOFs to any set of m DOFs. Potentially, it means that we can construct a process that can encode the information of a billion particles into a single particle.

TODO: add picture of diagonalization

Note that the above transformation is not possible in classical mechanics. One degree of freedom, topologically, is $\mathbb{R}^2$ while n DOFs are $\mathbb{R}^{2n}$, which are not topologically equivalent. The issue is that the topology of the Hilbert spaces in quantum mechanics is the one induced by the inner product and not the one induced by the observables, like in classical mechanics. Therefore the mathematical representation in quantum mechanics knows "too little" about the physical system it is supposed to describe.

Domain of operators

In the infinite dimensional case, the HellingerToeplitz theorem tells us that any Hermitian[4] operator O that is define on the whole space is bounded. Therefore any unbounded operator O will not be defined on the whole Hilbert space. Since variables like position and energy are unbounded, there will be some state ψ for which the energy is infinite, or the average position is undefined. This is clearly a problem, and it is instructive to understand where the problem comes from.

TODO: cannibalize the paper

TODO: add plots for the wave-functions

For example, consider the following wave function

$$\psi(x) = \sqrt{\frac{1}{\sqrt{\pi}} e^{-x^2}} \tag{3.44}$$
$$\rho_\psi(x) = \psi^\dagger(x)\psi(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}. \tag{3.45}$$

This is a Gaussian wave packet with expectation of zero and variance of $\frac{1}{2}$. Now consider the following wave function

$$\phi(x) = \sqrt{\frac{1}{\pi(x^2+1)}} \tag{3.46}$$
$$\rho_\phi(x) = \phi^\dagger(x)\phi(x) = \frac{1}{\pi(x^2+1)}. \tag{3.47}$$

Note $phi(x)$ goes to zero as $\frac{1}{x^2}$. The expectation, then, will converge and will converge to zero since the distribution is symmetric. However, the variance will diverge since $\lim_{x\to\infty} x^2 \frac{1}{\pi(x^2+1)} = \frac{1}{\pi}$.

[4] More precisely, symmetric. which means $\langle O\psi|\phi\rangle = \langle\psi|O\phi\rangle$.

Both of these will correspond to two vectors in the Hilbert space, meaning we are going to be able to find a unitary transformation that changes one to the other. In this case, we can do that by a change of variable. That is, we are going to look for a transformation $y = y(x)$ that transforms one wave function into the other. What we require is that the integral of one function over one region equals the integral of the second function on the second region. That is

$$\begin{aligned}
\int_0^{y(x)} \phi^\dagger(\hat{y})\phi(\hat{y})d\hat{y} &= \int_0^x \psi^\dagger(\hat{x})\psi(\hat{x})d\hat{x} \\
\int_0^{y(x)} \frac{1}{\pi(\hat{y}^2+1)}d\hat{y} &= \int_0^x \frac{1}{\sqrt{\pi}}e^{-x^2}d\hat{x} \\
\frac{\tan^{-1}(y(x))}{\pi} &= \frac{\operatorname{erf}(x)}{2} \\
y(x) &= \tan\left(\frac{\pi}{2}\operatorname{erf}(x)\right).
\end{aligned} \tag{3.48}$$

This change of variable, then, maps a state with finite expectation of position square to a state with infinite expectation. Changes of variable are unitary transformation on the Hilbert space, so in general a unitary transformation can map finite expectations to infinite expectations.

To be clear, the position in each reference frame corresponds to different observables. That is, X and Y are going to map to two different Hermitian operators. TODO finish the paragraph.

Moreover, in a Hilbert space we can always map one vector to another vector through a continuous unitary transformation, through continuous evolution. Mathematically, we can always rotate one vector on top of another. Physically, this means that we can construct a time evolution operator that oscillates between the two states. For example, suppose the evolution is such that the position chances in the following way

$$x(t) = \cos(\omega t)x_0 + \sin(\omega t)\tan\left(\frac{\pi}{2}\operatorname{erf}(x_0)\right). \tag{3.49}$$

This is a continuous transformation in t. For $t = 0$ we get the identity as $x(0) = x_0$. For $t = \frac{\pi}{2}$ we get $x(0) = \tan\left(\frac{\pi}{2}\operatorname{erf}(x_0)\right)$. Therefore $\psi(x)$ is the initial state, the evolution will oscillate between the two states, transforming finite expectation to infinite expectation and vice-versa in finite time, over and over.

Continuous spectra

In the finite dimensional case, we are used to associated possible values of an observable to states (i.e. the eigenstates) associated to that value. In the infinite dimensional case, we have operators that have a continuous spectra, like position or energy. There is a temptation to extend the previous scheme to the continuous one, adding eigenstates of continuous values. This is problematic, not just mathematically, but physically as well. Simply put, since we cannot prepare or measure a continuous quantity with infinite precision, these states are neither physically realizable nor measurable.

As we saw in the classical mechanics section, states as points in phase space (i.e. point particles) do not make sense in classical mechanics either. Volumes and areas define the geometry of phase space as these define the count of configurations per DOF and the count of

states. Hamiltonian mechanics is exactly the conservation of those areas and volumes. A single point, having no volume, can have no such notions. Classical point particles, then, should be understood as an infinitesimal region of phase space. A region so small that we do not care about its size, but still a region.

In quantum mechanics, talking about particles at a point makes even less sense. If we shrink the spread over position to zero, we are forced to stretch the spread over momentum to infinity. A uniform distribution over an infinite range makes even less sense than a distribution all concentrated at one point. Mathematically, a probability distribution over a single point is still a measure, a uniform probability distribution over an infinite range is not a measure. One way to see this, is that it does not satisfy countable additivity. We can imagine dividing the real line into countable intervals of equal size. Each should correspond to the same probability, and the sum of all the contributions should be one. If each interval corresponds to finite probability, the sum is infinite; if each interval corresponds to zero probability, the sum corresponds to zero probability as well. The solution would be to sacrifice countable additivity, so that each interval has zero probability while the total is one. Therefore, this would not satisfy the axioms of measure theory and probability theory.

Suppose we do want to go ahead and include distributions wholly concentrated at one point (i.e. delta function) in our state space. These are not square integrable function as they diverge, they are infinite, at one point. What is the inner product between such distribution and a standard state? It will be

$$\int \delta(x)\psi(x)dx = \psi(0). \tag{3.50}$$

This is called the sifting or sampling property of the delta function.

TODO:

, meaning that they do not have "proper" eigenstates. For example, for the position operator, the eigenvectors would correspond to the delta functions, which are not in the Hilbert space, but rather in the space of distributions. Therefore, there are no states with perfectly prepared position and the unitary transformation generated by position has no fixed points (i.e. equilibria, eigenstates).

Schwartz spaces

For multiple independent DOF of position and momentum, the requirement to have all polynomials of position and momentum with a finite expectation value recovers the Schwartz space, which is a dense subspace of the Hilbert space (i.e. any element of the Hilbert space can be understood as the limit of a sequence of Schwarz functions).

TODO: talk about the difference between Schwartz, Hilbert and Schwartz dual (distributions)

TODO: Does it make physical sense to have a mixed state over Schwartz where the probability coefficients converge as a polynomial.

Probability on a continuum

For any observable with any spectra, the probability measure can be recovered in the following way. Take a Borel set. Construct the projector onto that Borel set. For example, take X as the position operator. Take a Borel set A and take the indicator function 1_A and calculated $\langle\psi|1_A|\psi\rangle = \int_{\mathbb{R}} \psi^\dagger 1_A \psi dx = \int_A \psi^\dagger 1 \psi dx = \int_A \psi^\dagger \psi dx$.

Conjecture: whether finite expectations map to finite expectation if and only if the velocity is bound.

Self-adjoint vs Hermitian

In some spaces, self-adjoint is not equivalent to Hermitian. Consider the half real-line $[0, +\infty]$. Consider the momentum operator $\imath\hbar\partial_x$. The exponential $e^{-\lambda x}$ where $\lambda > 0$ is an eigenstate of momentum

$$\imath\hbar\partial_x e^{-\lambda x} = -\lambda\imath\hbar e^{-\lambda x} \tag{3.51}$$

with eigenvalue is $-\lambda\imath\hbar$ which is imaginary.

In the finite dimensional case, all self-adjoint are Hermitian.

Definition of Hermitian adjoint. Given O, the Hermitian adjoint $O^\dagger$ is such that:

$$\langle\psi|O\phi\rangle = \langle O^\dagger\psi|\phi\rangle \tag{3.52}$$

where $\psi, \phi \in D(O)$ (i.e. the operator is defined on the vectors).

Wavefunctions and equivalent classes

Schauder basis

TODO consider moving this into a bare minimum for quantum mechanics

Usually, one defines a basis B of a vector space V in linear algebra as a linearly independent subset of V that spans V. More explicitly

linear independence means that every *finite* subset $\{e_1, \ldots, e_n\} \subset B$ is linearly independent, i.e., $a_1e_1 + \ldots + a_ne_n = 0 \iff \forall 1 < i < n : a_i = 0$.

spanning means that for every vector $v \in V$ there exists a finite linear combination of some elements in B, s.t., $v = a_1e_1 + \ldots + a_ne_n$.

This definition is unpractical[5] for infinite-dimensional Hilbert spaces and the typical notion of basis differs from this. To differentiate we call the previous defintion that of a Hamel basis, while the following is called a Schauder basis.

A **Schauder basis** B of a separable Hilbert space $\mathcal{H}$ is a *sequence* e_n of vectors, s.t., for every vector $v \in \mathcal{H}$ there exists a *unique* sequence a_n such that

$$v = \sum_{n=1}^{\infty} a_n e_n \ . \tag{3.53}$$

Convergence is defined w.r.t the topology implicitly defined through the inner product on the Hilbert space.

Note that we had to explicitly define an *order* for the basis elements, since (3.53) may not converge unconditionally.

[5] I couldn't find any explicit example for a definition of a basis in the above sense for a separable Hilbert space. In fact, since it's existence is only guaranteed by the axiom of choice, I suspect it to be impossible to explicitly construct an example.

Physical states and well-defined moments

As we have seen before expectation values of observables are not always well-defined. Here we find another explicit example where a moment is not necessarily infinite, but depends on the order of basis.

Consider the state

$$\rho = \sum_{n=1}^{\infty} p_n |\psi_n\rangle\langle\psi_n| = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} |\psi_n\rangle\langle\psi_n| \tag{3.54}$$

and the observable

$$\mathcal{O} = \sum_{n=1}^{\infty} \mathcal{O}_n |\psi_n\rangle\langle\psi_n| = \sum_{n=1}^{\infty} (-1)^n n |\psi_n\rangle\langle\psi_n| \tag{3.55}$$

with some domain $\mathcal{D}_{\mathcal{O}}$, which makes $\mathcal{O}$ into a self-adjoint operator. Both of them are defined w.r.t. some ONB $\{\psi_n\}_{n=1}^{\infty} \subset \mathcal{H}$.

While calculating the expectation value of $\mathcal{O}$ for ρ in our ONB, we encounter the harmonic series[6]

$$\langle\mathcal{O}\rangle_\rho := \operatorname{tr}(\rho\mathcal{O}) = \sum_{n=1}^{\infty} p_n \mathcal{O}_n = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{(-1)^n n}{n^2} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} = \frac{6}{\pi^2} \ln(2) \tag{3.56}$$

This series is *conditionally* convergent and according to Riemann's rearrangement theorem it can therefore take *any* value in $[-\infty, +\infty]$ or not converge at all, just by rearranging the order of its terms, which is equivalent to a reordering of the basis elements.

Example for pure states

One can construct a similar example with a pure state. Consider

$$\psi = \sum_{n=1}^{\infty} \psi_n |\psi_n\rangle = \frac{\sqrt{6}}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} |\psi_n\rangle \tag{3.57}$$

and then calculate the expectation value as

$$\langle\mathcal{O}\rangle_\psi := \langle\psi|\mathcal{O}|\psi\rangle = \sum_{n=1}^{\infty} |\psi_n|^2 \mathcal{O}_n = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{(-1)^n n}{n^2} = \frac{6}{\pi^2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} = \frac{6}{\pi^2} \ln(2)\,. \tag{3.58}$$

The result is the same and will depend on the order of the basis elements in the same way.

While it's obvious that such a state can never be found in the lab, we need a criterion to distinguish the physical from the unphysical states.

For a pure state we can consider the domain of the operator as an indicator of what states have well defined expectation values. Let's check if ψ could possibly be in the domain of $\mathcal{O}$.

[6] This is of course no coincidence, but by construction.

For that, calculate

$$\begin{aligned}\mathcal{O}|\psi\rangle &= \sum_{m=1}^{\infty}(-1)^m m|\psi_m\rangle\langle\psi_m|\frac{\sqrt{6}}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}|\psi_n\rangle \\ &= \frac{\sqrt{6}}{\pi}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(-1)^m\frac{m}{n}|\psi_m\rangle\langle\psi_m|\psi_n\rangle \\ &= \frac{\sqrt{6}}{\pi}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}(-1)^m\frac{m}{n}\delta_{mn}|\psi_m\rangle \\ &= \frac{\sqrt{6}}{\pi}\sum_{n=1}^{\infty}(-1)^n|\psi_n\rangle\end{aligned} \tag{3.59}$$

and notice that the result is not normalized

$$||\mathcal{O}\psi||^2 = \frac{6}{\pi^2}\sum_{n=1}^{\infty}|(-1)^n|^2 = \frac{6}{\pi^2}\sum_{n=1}^{\infty}1 = \infty \tag{3.60}$$

and we conclude that ψ can not be in any sensible domain of $\mathcal{O}$.

Operator domain and the second moment

The previous observation that the domain indicates what states have finite expectation values can be generalized. Consider any unbounded observables $\mathcal{O}:\mathcal{D}_{\mathcal{O}}\to L^2(\mathbb{R}^n)$ and let's assume a very conservative domain of

$$\mathcal{D}_{\mathcal{O}} = \mathcal{D}_{\max} := \left\{\psi\in L^2(\mathbb{R}^n)\mid \mathcal{O}\psi\in L^2(\mathbb{R}^n)\right\}\ , \tag{3.61}$$

i.e., the resulting function must at least be in $L^2(\mathbb{R}^n)$.

Then, by definition it must be true that

$$||\mathcal{O}\psi||^2 < \infty \tag{3.62}$$

which is equivalent to

$$||\mathcal{O}\psi||^2 = \langle\psi|\mathcal{O}^\dagger\mathcal{O}|\psi\rangle = \langle\psi|\mathcal{O}^2|\psi\rangle = \langle\mathcal{O}^2\rangle_\psi < \infty \tag{3.63}$$

which means that at least the second moment has to be well-defined.

Insight 3.64. *States in the domain of unbounded operator must have at least a well-defined second moment.*

3.7 WARNINGS

Every state (even if I restrict to Schwartz space) is an eigenstate of some projection, observable and unitary, because $|\psi\rangle\langle\psi|$ is always defined.

Not every projector, observable or unitary can be understood as having eigenstates. For example, projection on a Borel set of position.

Questions: * Can we see the position operator as the limit of a discretized position operator where the discretization is smaller and smaller? If we did the same with momentum, would the commutator between them become $\imath\hbar$.

(?) Generalize this to arbitrary dimensions

(?) Random things to look at to see whether they are helpful * Khler manifold, interplay of symplectic structure with metric tensor * Is anything of the old arguments salvageable? * See if there is anything we can get from GPT or other reconstructions

Part IV

Appendix

Appendix A

Reference sheets for math and physics

A.1 Set theory

	Name	Meaning
$A = \{1, 2, 3\}$	set	a collection of elements
$\mathbb{N} = \{0, 1, 2, ...\}$	natural numbers	the set of numbers one uses to count
$\mathbb{Z} = \{.., -1, 0, 1, ..\}$	integers	the set of all whole numbers
$\mathbb{Q}$	rationals	the set of all fractions
$\mathbb{R}$	reals	the set of numbers with infinite precision
$\mathbb{C}$	complex	the set of numbers that represent a two dimensional vector or rotation
$a \in A$	in	whether the element a is contained in A
$A \subseteq B$	subset	a set that only contains elements of the other set
$A \subset B$	proper subset	a set that only contains elements of the other set but not all of them; it is a subset but is not the same set
$A \supseteq B$	superset	a set that contains all elements of the other set
$A \supset B$	proper superset	a set that contains all elements of the other set but not just them; it is a superset but is not the same set
$A \cup B$	union	the set of all elements contained in either sets
$A \cap B$	intersection	the set of all elements contained in both sets
$A \setminus B$	subtraction	the set of elements in A that are not in B
A^C	complement	the set of all elements that are not in A it is equal to $A \setminus U$ where U is the set of all elements, which depends on context
$A \times B$	Cartesian product	the set of all ordered pairs (a, b) with $a \in A$ and $b \in B$
2^A	power set	the set of all possible subsets of A

	Name	Meaning
$f : A \to B$	function	a map that for every element A returns an element of B
	injective function	a function that every distinct element of A map that for every element A returns an element of B
B^A		the set of all possible functions $f : A \to B$
$C(A, B)$		the set of all continuous functions $f : A \to B$

Credits

Created by:	Gabriele Carcassi
Written by:	Gabriele Carcassi and Christine A. Aidala
Subject-matter advisors (Math):	Mark Greenfield (Ch. II.1,II.2,II.3)
review prompted significant technical changes	
Additional subject-matter advisors (Math):	Daniel Burns, Alejandro Uribe, Alexander
review prompted significant technical improvements	Wilce (Ch. II.4)
(Phil):	Josh Hunt (Ch. II.1)
Subject-matter reviewers (Math):	Sharif Velasquez (Ch. II.1), Bart Westra (Ch.
review prompted technical fixes	II.3), Matt Insall, Junde Song (Ch. II.4)
Diagrams and figures:	Matteo Carcassi (Ch. I.1), Saja Gherri (Ch.
contributed one or more	II.1,II.2), Tobias Thrien (Ch. II.4)
Test readers:	Chami Amarasinghe, Andre Antoine, Hamza
reviewed a full chapter or more	Farooq, Alina Garcia, Saja Gherri, Uriah Israel, Micah Johnson, Sean Kelly, Dan McCusker, Pietro Monticone, Everardo Olide, Artem Omelchenko, Robert Rozite, Tobias Thrien
Additional test readers:	Josce Kooistra, Armin Nikkhah Shirazi, Ayla
review prompted corrections and clarifications	Rodriguez, Alex Takla, Allan Vanzandt