DRAFT
From physical principles to classical and quantum Hamiltonian and Lagrangian particle mechanics

Gabriele Carcassi, Christine A. Aidala, David J. Baker, Lydia Bieri
University of Michigan, Ann Arbor, MI 48109
(Dated: January 24, 2017)

This manuscript is a work in progress. Ideas are constantly reshaped to find more precise and elegant arguments. It is provided as is to stimulate discussion. **Make sure you have the latest version from http://assumptionsofphysics.org**

The aim of this work is to show that particle mechanics, both classical and quantum, Hamiltonian and Lagrangian, can be derived from few simple physical assumptions. Assuming deterministic and reversible time evolution will give us a dynamical system whose set of states form a topological space and whose law of evolution is a self-homeomorphism. Assuming the system is infinitesimally reducible, that giving the state and the dynamics of the whole system is equivalent to giving the state and the dynamics of its infinitesimal parts, will give us a classical Hamiltonian system. Assuming the system is irreducible, that giving the state and the dynamics of the whole system tells us nothing about the state and the dynamics of its substructure, will give us a quantum Hamiltonian system. Assuming kinematic equivalence, that studying trajectories is equivalent to studying state evolution, will give us Lagrangian mechanics and limit the form of the Hamiltonian/Lagrangian to the one with scalar and vector potential forces.

I. INTRODUCTION

Status:

- Section VII: Needs review.
- Section VIII: Needs review.
- Section IX: Needs review.

The past century has seen a resounding success in the use of mathematics in fundamental physics. In many cases, mathematical ideas were what allowed progress in theoretical physics and ultimately led to experimental discoveries. While these ideas are indeed useful—anything that can advance our understanding is welcome—they cannot, however, be the whole story.

The issue is that mathematical structures are not enough to characterize what physical system is being studied. For example, the same mathematical framework for linear circuits can be used to describe electric, hydraulic, thermal or mechanical systems [1–3]. The mathematical treatment is the same because it captures aspects of the description that are common to all cases; that indeed is its power. But for the very same reason, we cannot infer what system is being described if all we are given is the mathematical model. In the jargon of linguistics/computer science/philosophy [4, 5]: math captures the syntax (the relationships between the objects) but does not capture the semantics (the meaning of each object).

The real trouble comes when the mathematical description essentially becomes the foundation of a physical theory. While some branches of physics, such as Newtonian mechanics, thermodynamics and special relativity, are founded on physical laws or assumptions, others, such as Hamiltonian, Lagrangian and quantum mechanics, start by setting their mathematical structure. Nothing tells us why there should be a Lagrangian function whose path integral is minimized during the motion, or when a state is described by conjugate pairs (q^i, p_i) instead of a vector in a complex inner product space. Those are taken as given.

This means that if we have a system in front of us, we do not have a rigorous way to conceptually establish whether the system is Hamiltonian or not, whether it is classical or quantum. We have some heuristics we can apply, but, at the end the day, a system is Hamiltonian because it follows the Hamiltonian framework.[1]

Unsatisfied with this situation, we started creating a sort of math-to-physics dictionary so that each mathematical concept could be clearly associated to a crisp physical concept. As we pushed further back toward the basic mathematical definitions, we ended up with physical concepts of a more general nature and found that, in some cases, a single physical idea could have more than one mathematical consequence. That is: by putting physics at the center, we could achieve a unification of ideas that was not evident in the math alone. The unexpected success of the approach led us to think that, perhaps, the picture we were developing was a better framework to understand and link different areas of basic science and mathematics. This work is the systematized result of this endeavor.

The main goal is to show that it is possible to derive classical/quantum Hamiltonian/Lagrangian particle mechanics starting from few physical assumptions. We

[1] The lack of a properly defined, and uniquely accepted, semantics is of course very evident in quantum mechanics with its numerous interpretations.

do *not* claim that we have perfectly achieved this result, with no room for improvement. We *do* claim that the whole architectural framework is solid enough to be convinced that not only is this indeed possible, but it allows a more unified picture that facilitates intuitive connections within and across different areas of physics and mathematics. It enables a consistent narrative across several domains of knowledge.

Rederiving classical, quantum and (to some extent) relativistic mechanics within the same work is therefore crucial to show that the concepts we develop are truly general, that they can be applied to many different cases. If we could derive classical Hamiltonian mechanics on assumptions that are inapplicable to the quantum case, what good would it do? Quantum systems are Hamiltonian systems, so clearly we wouldn't have understood what makes a system Hamiltonian. This breadth of applicability is what makes the work valuable.

As a brief overview, there are four assumptions we will consider. Assuming deterministic and reversible evolution gives us a dynamical system: states and a law of evolution. Additionally assuming infinitesimal reducibility, studying the whole system is equivalent to studying its infinitesimal parts, gives us Hamiltonian mechanics. Further assuming kinematic equivalence, trajectories in space tell us everything about the state of the system, gives us Lagrangian mechanics and constrains the motion to massive particles under scalar/vector potential forces. Relativistic motion arises when studying time dependent laws of evolution with no further assumptions. Alternatively, as the second assumption we can instead take irreducibility, studying the whole system tells us nothing about its parts, which leads to quantum (Hamiltonian) mechanics.

As a more detailed overview: in section III we describe in general what it means to study a physical system and how the assumption of determinism and reversibility plays into it. We develop an abstract conceptual model that we use when discussing the later assumptions.

In section IV we show how the physical concept of physical distinguishability leads to the mathematical concept of topological spaces. Deterministic and reversible evolution is an invertible continuous map (i.e. a homeomorphism) as it needs to preserve what is physically distinguishable.

In section V we introduce the notion of a classical material, one for which we assume infinitesimal reducibility. The state of an arbitrary amount of material is a distribution over the states of the infinitesimal parts, the classical particles. As such a distribution must be invariant under the arbitrary choice of units, we see that for each variable q^i that defines a unit there is a corresponding variable k^i defined on the inverse of that unit. Classical particle states can be identified by conjugate pairs of state variables (q^i, p_i) (i.e. a point in the cotangent bundle $T^*\mathcal{Q}$) and the number of possible initial conditions for each pair is given by $dq^i \wedge dp_i$ (i.e. the canonical symplectic form ω). During evolution, the distribution (and marginal distributions) must be mapped point by point which leads to Hamiltonian mechanics (i.e. a symplectomorphism on $T^*\mathcal{Q}$).

In section VI we show that a relativistic version of Hamilton's equations arises naturally when we consider time dependent laws of evolution for distributions of classical material. This also gives the notion of deterministic and reversible evolution a characterization that is independent of any transformation that mixes time and state variables (q^i, p_i). Moreover it gives us a classical equivalent for anti-particle states.

In section VII we introduce the assumption of kinematic equivalence, studying the motion in space is equivalent to studying the evolution of the state. Under that assumption there is a link between the state variables (q^i, p_i) and the kinematic variables (x^i, u^i) (i.e. position and velocity). This link allows us to express the number of possible initial conditions for a degree of freedom (d.o.f.) in terms of dx^i and du^i, leading to the metric tensor g. Such an expression also allows us to constrain the Hamiltonian to the one for massive particles under scalar/vector potential forces and an equivalent Lagrangian formulation.

In section VIII we introduce some ideas in classical mechanics to more gently transition to quantum mechanics. We see how distributions over phase space already have features that are typically associated with quantum systems and we will show how to construct a classical uncertainty principle that helps to build intuition used in the later section.

In section IX we introduce the notion of an irreducible material, one for which studying the whole distribution does not tell us anything about the motion of its infinitesimal parts, the fragments. A unit amount of this material will correspond to a quantum particle. As this means that the states of such a material are invariant under permutations of the configurations of the fragments, we show that the state space of an irreducible material is a complex vector space. These states can still be compared to each other in a way that gives the state space an inner product. Such comparisons will be preserved by deterministic and reversible evolution which leads to Schroedinger's equation (i.e. a unitary transformation).

In section X we present a few higher level remarks about the work.

II. ORGANIZATION AND STYLE

The overall layout of this work is a balance between the need to show that the derivation follows, which requires starting from basic principles and proceeding using formal arguments, and the need to give an intuitive understanding, which often requires starting from specific cases and then generalizing.

The deductive aspect is formally organized into the following:

Assumptions: these characterize the physical system

we are studying and constitute the premise of our discussion. A **rationale** follows each assumption, which uses physical and sometimes philosophical arguments to motivate why (or not) such an assumption makes sense in a particular case.

Propositions: these capture the properties of our physical objects into mathematical language. A **justification** follows each proposition to show the necessity of such a characterization. While it is a formal argument, it is not a strictly mathematical proof as it contains physical arguments.

Theorems: these are pure mathematical statements that are used to simplify propositions. A mathematical **proof** follows each theorem. No mathematical breakthrough should be expected as we mostly use well known results from different fields. Theorems are grouped in the appendix.

As these sections are clearly demarcated, they allow us to make sure that each justification only references past propositions and avoid circular arguments. It also makes it easier to read the work at the desired level of detail. Interspersed between each step of the more formal treatment, there are numerous discussions that conceptually explain motivations and results in a more approachable way. These can use more concrete and intuitive examples which may, in some cases, make forward references. As both the formal and more intuitive parts need to cover the same material, this will inevitably lead to some repetition.

The work uses ideas and tools from a broad range of mathematical and scientific subjects: topology [6], measure theory (ref?), differential geometry (symplectic and Riemannian) [7], group theory and vector spaces [8, 9], statistics (ref?), information theory (ref?), Hamiltonian mechanics, Lagrangian mechanics [10], quantum mechanics (ref?), thermodynamics (ref?), special relativity (ref?) and so on.[2] Using ideas from all these areas is necessary and valuable, but it does create a number of problems.

The first one is terminology and prerequisite knowledge. Since we touch numerous areas, we cannot expect all readers to be equally comfortable with all the technical terms. Yet we cannot introduce all of them either. The compromise we adopted is to keep the informal discussion more accessible while keeping the formal derivation more precise, without needing more than the basic definitions and few major results within each specific area. This may leave an expert in a particular field not fully satisfied but, as stated in the introduction, our goal is to lay out the general picture.

The second problem is notation. As each field has developed its own set of conventions, the common usage is now an empty set. We ended up with a compromise that hopefully feels natural enough to a broad audience of physicists, mathematical physicists and philosophers of physics. Within formal parts we may use notation and terminology more specific to the appropriate area of math.

We have tested the prose on a number of colleagues and students in physics, philosophy and mathematics to confirm its accessibility at different levels and believe we have struck a reasonable compromise.

III. ON STUDYING A PHYSICAL SYSTEM

Our first task is to develop a conceptual model that applies to all realms of physics we'll be considering: classical, statistical and, later, quantum mechanics. We will take the standard picture of system plus environment and extend it to differentiate between the state of the system (i.e. the aspects under study) from the unstated part of the system (i.e. the aspects missing from our description).

We will assume that the state evolves according to a deterministic and reversible law (i.e. for each present state there is one and only one future state). While the unstated part does not influence the state evolution, we'll see how it constrains what states are available and what type of description can or cannot be given to the system.

III.A. States and their evolution

We start by fixing a *physical system*, meaning something we can interact with and perform measurements on (e.g. a planet, a fluid, ...). We call *environment* everything else. We set what particular aspect we want to study (e.g. the motion around a star, the flow in a pipe, ...). We call *state* a physically distinguishable configuration of the aspect under study at a particular time (e.g. position/momentum of center of mass, velocity field, ...). Since the state does not, in general, exhaust the description of the system, a part remains *unstated*, and as such we'll call it, for lack of a better word (e.g. the chemical composition, the motion of each of its molecules, ...). Note that the environment plays an essential role here as it's what allows us to define two states as physically distinguishable: we can find an external process (i.e. part of the environment) whose outcome changes depending on the different state. As such processes are what we can use to perform measurements, we consider the experimental apparatus (and us performing measurements) independent of the system, part of the environment.[3]

[2] The references given provide the reader a list of resources consistent with the development of this work.

[3] This is true even in the case of general relativity, where we can imagine multiple researchers on small spaceships collecting data without greatly influencing the motion of stars and planets. The

In this context, we call the evolution of the system *deterministic* if the state at a given time uniquely identifies states at future times, and *reversible* if it uniquely identifies states at past times. While this is a common enough definition, we need to be clear how this applies to the unstated part. Note that the concept of determinism outlined here is context dependent because the state itself represents only the part of the system that we choose to (or can) describe. In this sense, the unstated part is *always* non-deterministic (and non-reversible) in the sense that the state of the system does not determine its evolution. For example, suppose we define the state as position and momentum of the center of mass of a cannonball. Suppose that the evolution is deterministic on that state. What does it tell us about the cannonball temperature, or about the motion of each of its atoms? Nothing. In this sense, the unstated part is non-deterministic and non-reversible. Could we extend the state and the laws of evolution to account for temperature? Yes, but that would be a different evolution defined on a different state. Does it mean the unstated part is always evolving chaotically? Not at all. The temperature may remain constant throughout the motion of the cannonball. Yet we wouldn't know, since we are not studying it: the evolution is deterministic and/or reversible only as far as the state is concerned. With this in mind, we will restrict ourselves to the cases where the following is valid:

Assumption I (Determinism and reversibility)**.** *The state of the physical system under study undergoes deterministic and reversible evolution.*

Rationale. As it is an assumption, we first need to discuss when it is valid. More specifically, we need to understand that the non-deterministic/non-reversible evolution of the unstated part plays as much of a fundamental role as the deterministic/reversible evolution of the state. In fact, the non-deterministic part contributes in determining what states are available to the system.

Suppose we study the motion of a cannonball; its state under gravitational and (inertial) inertial forces will be properly described by the position and momentum of the center of mass. While light and air molecules may scatter off its surface unpredictably, its trajectory is not greatly affected as it is a massive rigid body. Suppose we study the motion of a small particle, small enough that the random scattering does influence the trajectory and it undergoes Brownian motion: its state will be a probability distribution for position and momentum of the center of mass. Gravitational and inertial forces have not changed, yet the states have changed from "pure" to statistical ensembles. In other words, the set of states must be closed under both the deterministic evolution of the state and the non-deterministic evolution of the unstated part. If the Brownian motion is not negligible, we do not end in a well defined position/momentum pair, even if we start from one.

A similar more drastic effect: consider a book and its motion under gravitation and inertial forces, its state being the position and momentum of the center of mass. As we increase the temperature of the air around the book, its motion remains unaffected until, at some point, the book burns. Clearly, the non-deterministic evolution has pushed one of the states outside the set of states, to the point that the system is no longer recognizable.

As we have hinted, sometimes the state is identified by a distribution (either statistical or actual). Even in this case, the state can be deterministic and reversible. That is, given the distribution at one time we can determine the distribution at future times. The shape and the parameters of the distribution can be deterministic, even if the evolution of the parts are not as they fall within the unstated part. Note that we *cannot* assume trajectories and states are always defined for the unstated part, as this includes also the unknown unknowns. We will return to this aspect when discussing quantum systems.

It should also be clear that what constitutes state and unstated part does not depend only on the system under study, but also on the processes we are considering. In some circumstances, the chemical composition of a fluid may be relevant, in others it may not. The choices of environment, state and unstated part are not independent from each other. By choosing a particular set of states, we are not only saying that the state evolution is well approximated by a deterministic/reversible map from initial to final state, but we are also saying that the non-deterministic/non-reversible evolution of the unstated part does not change the nature of the system, and processes that do not satisfy these conditions are not under consideration.

As with all assumptions, we should also ask whether it is necessary. That is, could we define a set of physically distinguishable states and yet have no deterministic and reversible processes defined on them? The claim is that this assumption is indeed needed, as without it we cannot properly define states or write useful physics laws. We can provide different arguments that point in the same direction.

First, to be able to identify the system, we must be able to tell it apart from anything else. Intuitively, we can distinguish between two chairs because we can move the first to another room and sit on it without having touched the second. We can manipulate the state of the first system without affecting the second, and vice-versa. So, to identify a system it has to be sufficiently isolated from everything else. This means that the system future

case where the physical system is the whole universe and there is no environment presents practical problems and conceptual challenges that the current physical theories do not seem to be equipped to address, and therefore will be absent from our discussion. For example, what physical device can we use to store and process the state of the whole universe to make predictions and compare? How do we define physically distinguishable? Do the physical laws determine which measurements we are going to make and does that limit what is actually distinguishable?

and past states are with good approximation determined only by its own state: the state undergoes deterministic and reversible evolution.

Second, the aim of physics is to write laws that can be used to make predictions that can be validated experimentally. If I drop an anvil from a tower, it will accelerate at 9.81 m/s^2; if I want the anvil to reach the ground at x m/s I have to drop it from y m. To the extent that we want to make predictions in time, we need to have a correspondence between initial and final states.

Third, operationally we must reliably prepare and measure states. That is, we need a process for which the input settings of our preparing device determine the outgoing state of the system; and a process for which the incoming state of the system can be reconstructed by the output of the measuring device. That is, our system must, at least in some cases, be able to participate in a deterministic and reversible process with the preparing and measuring device. Without it we wouldn't be able to calibrate our experimental apparatus.

Fourth, to be able to ascribe a property to a system we need to claim that, at least for a finite interval of time, the system either held or did not hold such property. That is, there is a deterministic and reversible process for that finite period of time for such property.

This link between state definition and deterministic processes should not be too surprising as the state, in the context of control theory and dynamical systems theory, is often defined as *the set of variables needed to determine the future evolution of the system.*(ref?) As we saw before, this applies also to statistical processes: the distribution (the ensemble) as a whole can indeed be calculated, measured and prepared. We can also describe the evolution of each element of the distribution provided that: we have a way to isolate it and study it under deterministic and reversible motion (so that we can define microstates); the non-deterministic motion does not alter the system (the set of microstates is preserved by the evolution).

As with many assumptions, we should stress that it's an idealization: it can never be completely achieved in practice. A system can be prepared or measured up to a certain level of precision. Perfect determinism and isolation of a system is impossible both practically (e.g. blackbody radiation, gravity, ...) and conceptually (e.g. if the system is perfectly isolated, we cannot interact with it: how can it be physically distinguished?). It's a simplifying assumption that can only be taken if the environment and the internal dynamics of the system interact in such a way that they little affect and are little affected by the aspect we are studying. As we saw before, for example, assuming that the state consists of the position and momentum of the center of mass requires assuming that the Brownian motion of the body is negligible.

Yet, this is a fundamental assumption in the sense that it is needed. If a particular set of states does not satisfy deterministic and reversible evolution under certain conditions, what we do is to keep at it until we find a set that does. That is, we work to restore the assumption. Finding new sets of states with new laws of evolution is, in fact, what leads to new physics. We therefore call *fundamental model of physics* the triad of state, unstated part and environment, with the assumption that the state may undergo deterministic and reversible evolution, and the unstated part undergoes non-deterministic non-reversible evolution that does not alter the set of states. □

IV. STATES AND STATE SPACE

We now proceed to characterize states and physical distinguishability in more precise terms so that we can capture their description mathematically. What we'll see is that the outcomes of all physical processes that can be used to gain information about the system, that can be used to perform a measurement, induce a topology on the set of states that is at least Hausdorff. That is, physical distinguishability is mathematically captured by topological distinguishability. Deterministic and reversible evolution will then preserve the topology, and they will be mathematically captured by self-homeomorphisms.

We'll focus on state spaces that can be described by a set of independent state variables, either discrete or continuous, and see what can be said in general on the evolution of state variables.

IV.A. States and topology

As the term "measurement" has become particularly loaded, let's first characterize what we mean by physical distinguishability in our context.

Consider the motion of a cannonball under inertial and gravitational forces. Light will scatter off of it; as it lands the ground will be deformed and the impact will make the temperature slightly rise. Those external physical processes, which happen no matter what we do, can be used to distinguish the motion of the cannonball as their outcomes are correlated. Therefore we can learn the position by looking at the reflected light, and learn the final kinetic energy by looking at the deformation of the ground. In principle, any external process that has a correlation with the states under study can be used to perform a measurement, and any measurement is based on such a process. That is, for us a measurement is simply a physical process that we can use to distinguish states. Setting up an experiment means choosing a particular process with desired outcomes and forcing the system under study to interact with it one or more times. After that, there is no special role played by the "observer" in making outcomes come about.

This external process may be quite complicated: when a particle enters a calorimeter, a shower of particles is produced, photons are captured and are directed to photomultipliers, a current is read out, the current is then digitized, and so on. Some processes interfere with the

system, they affect its dynamics, and others are destructive, the system no longer can be described by the original set of states. A tracking chamber is an example of the first (the magnetic field curves the motion of a charged particle); burning a substance to determine its caloric content is an example of the latter. Therefore intimate knowledge of the process is always needed to ensure that one makes the proper link between outcomes and the *original* states, and properly accounts for systematic uncertainties that would skew that link.

Repeatability is also fundamental. First, to make sure the process is indeed correlated to the states. Second, because "a single take does not a measurement make". One has to gather enough statistics. Note that the number of takes influences the outcomes: with greater statistics the precision and number of distinguishable cases increases. Therefore the processes, as we defined them, may require repeated interactions with similarly prepared states. They may even be a combination of different kinds of interactions that, taken all together, create a set of distinguishable outcomes. But in the end, however complicated it is, the conceptual model remains the same: each process has a set of possible outcomes, and each possible outcome will be associated with a set of states consistent with that outcome. For example, if the cannonball deformed the ground by this amount then its kinetic energy at impact was within this range; if the electron follows a certain path, then its state is among the ones that have spin $+1/2$.

Note that the outcome of the process (e.g. the electron followed a particular path) is conceptually distinct from the set of states compatible with it (the states associated with the path) and the way we label them (spin=$+1/2$). Yet, we'll conflate the concepts as this makes the language closer to experimental physics ("the outcome of the measurement is spin=$+1/2$"). Context will allow us to recover the distinction.

However precise our measurements are, we can only gather a finite amount of statistics and each outcome is expressed by a finite set of digits; therefore, the set of outcomes is countable.[4] Note that different outcomes for the same process can overlap. For example, 4.12 ± 0.05 cm and 4.13 ± 0.05 cm are both legitimate possible outputs of the same measurement device. But since all states must be distinguishable, given two arbitrary states there must be a process precise enough to tell them apart. That is, the potential outcomes associated with the two states do not overlap. For example, 4.12 ± 0.0005 cm and 4.13 ± 0.0005 cm do not overlap anymore.

We can also conceptually combine two different processes into a single one.[5] That is, having a way to measure quantity x and a way to measure quantity y gives us a way to measure the combination (x, y). For ensembles, if we can measure the marginal distribution $\rho_x(x)$ and the marginal distribution $\rho_y(y)$, we know the joint distribution $\rho(x, y)$ has to be compatible with both. Note, though, that this does not provide a way to fully measure $\rho(x, y)$ as we know nothing about the correlation between the two variables.[6] Formally, the states compatible with the outcomes of the combined process will be the intersections of the states compatible with each pair of outcomes of the original processes.

We can also coarsen a single process.[7] That is, having a way to measure (x, y) gives us a way to measure x alone (or any $f(x, y)$). Formally, the outcomes of the coarsened process are given by performing the union of some outcomes of the original process.

This model maps very naturally to a topological space. The states are elements of a set and the physical outcomes provide a topology on that set. The fact that two elements of the set can be distinguished requires the space to be Hausdorff.

Proposition IV.1. The state space $\mathcal{S}$ of a physical system is a Hausdorff topological space.

Justification. We claim $\mathcal{S}$ is a set. Each state is well defined as it is physically distinguishable. The collection of all possible states forms a set.

We claim $\mathcal{S}$ has a topology T. Consider the set of all possible physical outcomes associated with all physical processes. Each possible outcome is associated with a set of states that are compatible with that outcome. Let T be the set of all sets associated with all physical outcomes. $\mathcal{S} \in \mathsf{T}$ and is associated with the outcome "the system exists". $\varnothing \in \mathsf{T}$ and is associated with the outcome "the system doesn't exist". Let $V_1, V_2 \in \mathsf{T}$. Then, by definition, there exists a process P_1 that admits V_1 as an outcome and a process P_2 that admits V_2 as an outcome. Consider the process P that combines the outcomes of P_1 and P_2 with a logical AND. This always exists physically as we can prepare the same state multiple times and let it interact with each process separately. P will have a possible outcome V corresponding to the case where P_1 gave outcome V_1 and P_2 gave outcome V_2. The states compatible with V must be in both V_1 and V_2, that is $V = V_1 \cap V_2$. Therefore T is closed under intersection. Let $V_1, V_2 \in \mathsf{T}$. If they are physically distinguishable, then there exists a physical process P that admits both as outcomes. Given P, we can always construct the process P_0 that combines V_1, V_2 with a logical OR into a single outcome V, by "forgetting" which of the two was given. The states compatible with V must be in either V_1 or V_2, that is $V = V_1 \cup V_2$. Therefore T is closed under union. T is a topology by definition.

[4] The information provided by the process as measured by Shannon's entropy[11, 12] is finite.

[5] That is, take the logical AND between outcomes of different processes.

[6] The quantum case is similar: we measure marginal distributions and rule out states that are incompatible with them.

[7] That is, take the logical OR between outcomes of the same process.

We claim that T is Hausdorff. Let $\mathscr{s}_1, \mathscr{s}_2 \in \mathcal{S}$. As states are physically distinguishable, there must exist a physical process with two possible outcomes $V_1, V_2 \in \mathsf{T}$ for which $s_1 \in V_1, s_2 \in V_2, V_1 \cap V_2 = \varnothing$. $\mathcal{S}$ is Hausdorff by definition. □

Note that the arguments that led to the topological space had nothing to do with states per se, just that they are physically distinguishable. States, though, are not the only objects with that property. In fact, any element of a set of physical objects (e.g. forces, physical properties such as mass or charge, time) needs to be physically distinguishable to be well defined. We can generalize the above justification: any set of physically distinguishable elements is a topological space.

Proposition IV.2. Any set $\mathcal{S}$ of physically distinguishable elements is a Hausdorff topological space.

Justification. Same justification as in IV.1 with "state" replaced by "element" of the set $\mathcal{S}$. □

With our state space defined, deterministic and reversible evolution corresponds to a bijective map between initial and final states. But that's not enough. A physical process that can distinguish final states can be used, together with the evolution, to distinguish initial states. We prepare the initial state, let it evolve deterministically, measure the final state and use reversibility to convert the measure to one on the initial state. Therefore, not only the evolution is a bijection, but it also maps outcomes to outcomes. That is: the topology is mapped and preserved by the deterministic and reversible evolution because physical distinguishability must remain unchanged. The evolution is then a self-homeomorphism on the state space.

Proposition IV.3. A deterministic and reversible evolution map is a self-homeomorphism on the state space $\mathcal{T}_{\Delta t} : \mathcal{S} \to \mathcal{S}$.

Justification. We claim that $\mathcal{T}_{\Delta t}$ exists. The system is deterministic: given an initial state $\mathscr{s} \in \mathcal{S}$ there exists a well defined final state $\mathcal{T}_{\Delta t}(\mathscr{s}) = \hat{\mathscr{s}} \in \mathcal{S}$.

We claim $\mathcal{T}_{\Delta t}$ is continuous. Let $U \subseteq \mathcal{S}$ represent an outcome of a process P on the final states. U is an open set in the (final) state space topology by definition. Consider the process P_0 that first evolves the initial states and then distinguishes the final state with P. P_0 is a process that distinguishes initial states. The set of initial states compatible with U are $\mathcal{T}_{\Delta t}^{-1}(U)$. $\mathcal{T}_{\Delta t}^{-1}(U)$ is an open set in the (initial) state space topology by definition. $\mathcal{T}_{\Delta t}$ is a continuous map.

We claim that $\mathcal{T}_{\Delta t}$ is a bijection. The system is reversible: there exists a map $\mathcal{T}_{-\Delta t} : \mathcal{S} \to \mathcal{S}$ that returns the initial state given the final state. $\mathcal{T}_{-\Delta t} \circ \mathcal{T}_{\Delta t} = \mathcal{T}_{\Delta t} \circ \mathcal{T}_{-\Delta t} = \mathrm{id}_{\mathcal{S}}$ as mapping forward and then backward or backwards and then forward must return the original element. $\mathcal{T}_{\Delta t}$ admits $\mathcal{T}_{-\Delta t}$ as an inverse. $\mathcal{T}_{\Delta t}$ is a bijection.

We claim that $\mathcal{T}_{\Delta t}$ is a self-homeomorphism as it is a continuous bijection. □

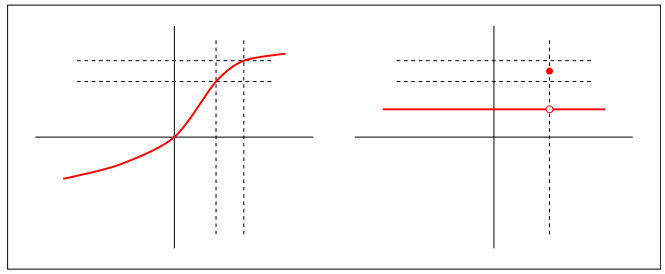

FIG. 1. A continuous function always maps finite precision knowledge in the vertical axis (i.e. an open set) to finite precision knowledge in the horizontal axis. A discontinuous function may map to infinite precision (i.e. a closed set). If infinite precision identification is ruled out (i.e. use of standard topology) then discontinuous functions must also be ruled out.

Again, we note that the arguments that lead to continuity had nothing to do with states per se, just that the relationship is between physically distinguishable objects. We can then generalize the above justification.

Proposition IV.4. A map $f : \mathcal{S}_1 \to \mathcal{S}_2$ between two sets of physically distinguishable elements $\mathcal{S}_1$ and $\mathcal{S}_2$ is a continuous map.

Justification. Same justification for continuity as in IV.3 with "initial states" and "final states" replaced by "elements" of $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively. □

The generality of this result explains why in physics one always assumes functions to be "well behaved". As the result was derived from our notion of physical distinguishability, this is not a matter of practical convenience. Suppose we were able to prepare a force field that was zero everywhere in space except at a single point. This gives us a way to tag a specific point. But it also allows us to create an outcome compatible with only a single point. A finite precision measurement of the force provides us infinite precision of space, which we ruled out.[8] This is what the math is telling us, that if we claim that outcomes can't distinguish isolated points (i.e. standard topology on $\mathbb{R}^n$) then neither can maps. It is physical consistency that limits us to continuous functions.

While functions with few discontinuities are useful and used in physics and engineering, they are employed for idealized cases (e.g. a signal change is fast enough, a charge distribution is small enough) that are often treated as a special case (e.g. propagation across material discontinuities). While this may be obvious and intuitive to the physicist, it may be troubling to the mathematician as the proper use of many mathematical techniques requires the inclusion of discontinuous functions. This is less of a problem than it would seem at first. Once we

[8] This assumes that we are able to position the probe perfectly, which we could do if we were able to manipulate forces at that precision.

made sure that the objects and their relationships are physically meaningful, we can extend our mathematical spaces for the purpose of math computations. Our physically meaningful continuous function can be expanded into a sum of discontinuous functions. One just has to be mindful of the extension and be wary that mathematical results that depend on such extension may or may not be physically meaningful.

IV.B. Manifolds and labeling states

To identify and name states one uses a set of quantities, typically numbers. For example, the orbital of an electron in a hydrogen atom is identified by the quantum numbers n, l, m and s. We call each of these quantities *state variables*. We call a *possibility* a possible value that can be taken by a state variable. We call a state variable *discrete* or *continuous* if the possibilities are integer or real numbers respectively. The topology used reflects the notion that integers can be measured perfectly (i.e. discrete topology) while real numbers can be measured up to a finite precision (i.e. standard topology on $\mathbb{R}$). From now on, we are going to consider state spaces whose states can be identified, at least within a region, by a finite set of discrete and continuous state variables (i.e. the state space is locally isomorphic to $\bigcup_{1\leq i\leq n} \mathbb{R}^{m_i}$).

We purposely use the term state variables instead of coordinates (even though that's what they are mathematically) as it would create confusion with space-time coordinates. We also avoid the term observable or measurable, as not all state variables may be directly physically tangible (e.g. conjugate momentum in a gauge theory). The only requirement for state variables is that they identify states. This means the possible values for state variables are physically distinguishable and are defined at equal time (since states are defined and mapped at a particular time).[9]

Proposition IV.5. A *state variable* is a continuous map $q : U \to \mathbb{L}$ where $U \subseteq \mathcal{S}$ and $\mathbb{L}$ is the space for the possible values. If $\mathbb{L} \cong \mathbb{Z}$ the variable is said *discrete*. If $\mathbb{L} \cong \mathbb{R}$ the variable is said *continuous*.

Justification. We claim $\mathbb{L}$ is a topological space. $\mathbb{L}$ is a set of physically distinguishable possibilities. $\mathbb{L}$ is a topological space because of IV.2.

We claim q is a continuous map. q is a map between two sets of physically distinguishable elements. q is continuous because of IV.4. □

When combining multiple state variables, it is important to understand how they relate to each other. Consider the orbital of an electron in a hydrogen atom, which is identified by the quantum numbers n, l, m and s. For each combination of n, l and m, the spin s can have two values. The choice of spin is independent from the rest. The choice of l, though, depends on the choice of n: for $n = 1$ only $l = 0$ is available; for $n = 2$ we can choose $l = 0$ or $l = 1$. The choices are not independent. That is: two or more variables are independent if there always exists a state for any possible combination, if the total number of states is the product of the possibilities of each variable. We call *state vector* a collection of independent state variables that fully identify a state.

Proposition IV.6. Two state variables $q_1 : U \to \mathbb{L}_1$ and $q_2 : U \to \mathbb{L}_2$ are said *independent* if $\exists \mathscr{s}$ such that $q_1(\mathscr{s}) = l_1, q_2(\mathscr{s}) = l_2 \forall l_1 \in q_1(U), l_2 \in q_2(U)$.

In general, the entire state space may not be identified by a predetermined set of independent state variables. Consider the state of a pool table, determined by the number of balls together with position and momentum of the center of mass: the number of state variables is not fixed as it depends on the number of balls. But deterministic and reversible evolution cannot take us from a different number of continuous independent state variables (i.e. there is no homeomorphism between $\mathbb{R}^n$ and $\mathbb{R}^m$). That is the number of balls cannot change under deterministic and reversible evolution. Therefore we can restrict ourselves to the case where the number of continuous independent state variables is constant without loss of generality. This means that, at least locally, the state space is always homeomorphic to $\mathbb{R}^n$, and is therefore a manifold.

Discrete variables do not present such problems, as any number of them can be flattened out in a single one (i.e. $\mathbb{Z}^n$ is homeomorphic to $\mathbb{Z}$). They will determine the number of connected components of the state space. Once we introduce a continuous parameter for time evolution, though, these become irrelevant as continuous time evolution requires continuous trajectories that cannot move states across disconnected components. This justifies the special interest in path connected manifolds, as this is where trajectories for deterministic and reversible continuous evolution live.

Proposition IV.7. Let $\mathscr{s} \in \mathcal{S}$ a state within a state space. The set of states $\mathcal{S}' \subseteq \mathcal{S}$ potentially reachable from $\mathscr{s}$ by deterministic and reversible continuous evolution is a path connected manifold of dimension equal to the number of independent continuous state variables necessary to identify it.

Justification. We claim $\mathcal{S}'$ is a manifold. Let $\mathscr{s} \in \mathcal{S}$. Let n be the number of independent continuous variables needed to identify $\mathscr{s}$. There exists a neighborhood U

[9] This is the main reason that a quantity like velocity is not a suitable state variable, as it is defined over an interval of time. Therefore velocity is always physically well defined but is not a state variable in general, while conjugate momentum is always a state variable but is not physically well defined by itself (it requires the vector potential to be specified as well). As we'll see later, when there exists a one to one map between velocity and conjugate momentum they can both be physically well defined state variables.

around $\mathscr{s}$ where we have $(q_1, ..., q_n) : U \to \mathbb{R}^n$. This map is a bijection as $\mathscr{s}$ is identified by those variables. $\mathcal{S}$ is homeomorphic to $\mathbb{R}^n$ around $\mathscr{s}$. Let $\mathcal{S}'$ be the set of all states potentially reachable by deterministic and reversible evolution from $\mathscr{s}$. Deterministic and reversible evolution is a homeomorphism between initial and final states. $\forall \hat{\mathscr{s}} \in \mathcal{S}'$ there exist a map $\mathcal{T}_{\Delta t} : \mathcal{S} \to \mathcal{S}$ such that $\hat{\mathscr{s}} = \mathcal{T}_{\Delta t}(\mathscr{s})$. $\hat{\mathscr{s}} \in \mathcal{T}_{\Delta t}(U)$ and $\mathcal{T}_{\Delta t}(U)$ is isomorphic to $\mathbb{R}^n$. $\mathcal{S}'$ equipped with the subspace topology is a topological space everywhere homeomorphic to $\mathbb{R}^n$. $\mathcal{S}'$ is a manifold of dimension n.

We claim $\mathcal{S}'$ is path connected. Let $\hat{\mathscr{s}} \in \mathcal{S}'$. There exist map $\lambda : [t_0, t_1] \to \mathcal{S}'$ where $\lambda(t_0) = \mathscr{s}$, $\lambda(t_1) = \hat{\mathscr{s}}$ and t_0 and t_1 are the initial and final time respectively. λ is a map between two physically distinguishable quantities, time and states. λ is continuous by IV.4. All $\hat{\mathscr{s}} \in \mathcal{S}'$ are path connected to $\mathscr{s}$. $\mathcal{S}'$ is path connected. □

IV.C. Evolution of state variables

To study time evolution, we need to describe how state variables change under deterministic and reversible evolution. There are two ways to do it, and we'll need both. The first approach is to *evolve* the state variables from the initial value $q^i(\mathscr{s})$ to the final value $q^i(\hat{\mathscr{s}})$. The result is a trajectory $q^i(t) = q^i(\lambda(t))$ which is especially useful when state variables correspond to physically meaningful quantities. For example, we track how the temperature or the pressure of an ideal gas changes. Evolved state variables, however, make it hard to find relationships that are invariant and common to all deterministic and reversible processes.

The second approach is to transport the state variables. The idea is to keep the connection to the initial state by labeling the future state by the original value of q^i, instead of by the future values. For example, the evolved state variable will not tell us the current value of pressure, but the one for the initial conditions. So we introduce a new set of variables for which $\hat{q}^i(\hat{\mathscr{s}}) = q^i(\mathscr{s})$, which we can always do as the evolution is deterministic and reversible. The hypersurfaces at constant $\hat{q}^i$ are the evolved hypersurfaces at constant of q^i (e.g. all the states that started with a particular value for pressure) which allows us to study how groups of states evolve in time. Given that the value of the transported state variables does not change during evolution, and that it is unique for each initial state, transported state variables also provide a way to label the trajectories themselves.

Evolved state variables are useful to write equations of motions, study how physical quantities change, form a physical picture of what happens. Transported state variables are useful to write invariants, study state space trajectories, form a geometric picture for the state space.[10]

Proposition IV.8. Let q^i be a set of state variables and $\mathcal{T}_{\Delta t}$ a deterministic and reversible evolution map. The evolved state variables are given by $q^i \circ \mathcal{T}_{\Delta t}$. The transported state variables are given by $q^i \circ \mathcal{T}_{-\Delta t}$.

One reason that transported variables will be useful is that, during deterministic and reversible evolution, we need to make sure that the number of initial and final possibilities across independent state variables remains the same. This is more easily done using transported variables as the actual values do not change. In the discrete case, this is easy to show.

Proposition IV.9. Let $U \subseteq \mathcal{S}$ a set of states fully identified by a set of n independent discrete state variables q^i. Let $\Delta q^i \equiv q^i(U)$ the range of possibilities of each variable. Then $\#(\Delta q^i) = \#(\Delta \hat{q}^i)\ \forall i$ and $\#(U) = \prod_{i=1}^{n} \#(\Delta q^i) = \prod_{i=1}^{n} \#(\Delta \hat{q}^i) = \#(\hat{U})$ where $\hat{U} = \mathcal{T}_{\Delta t}(U)$ and $\#$ denotes the number of elements in the given set.

Justification. We claim $\#(\Delta q^i) = \#(\Delta \hat{q}^i)$. For each $\mathscr{s} \in U$, let $\hat{\mathscr{s}} = \mathcal{T}_{\Delta t}(\mathscr{s})$. We have $\hat{q}^i(\hat{\mathscr{s}}) = q^i(\mathscr{s})$. Therefore $\Delta q^i = q^i(U) = \hat{q}^i(\hat{U}) = \Delta \hat{q}^i$. $\#(\Delta q^i) = \#(\Delta \hat{q}^i)$.

We claim $\#(U) = \prod_{i=1}^{n} \#(\Delta q^i)$. As q^i are independent variables, the states are the Cartesian product of the possibilities of each variable. □

What happens is that, because the variables are independent, the total number of states is the product of the number of possibilities for each variable. If Δq^1 and Δq^2 are ranges of possibilities for two independent variables, the total number of possibilities is $\#(\Delta q^1)\#(\Delta q^2)$. By construction, the set Δq^i of possibilities for each independent variable is the same as the set $\Delta \hat{q}^i$ of possibilities for the transported variable. Also, the transported variables remain independent. Therefore the relationship $\#(\Delta \hat{q}^1)\#(\Delta \hat{q}^2) = \#(\Delta q^1)\#(\Delta q^2)$ is valid throughout the evolution.

We'll see that very similar relationships are what define Hamiltonian and Lagrangian mechanics. But recovering them in the continuous case is not as straightforward. Consider the map $q' = aq$ with $0 < a < 1$. At first glance, it's a bijective continuous map so we may think it can represent a deterministic and reversible evolution. Yet, $\Delta q \equiv [-b, b] \supset \Delta q'$: a set of states is mapped to a proper subset (i.e. to fewer states) which does not make sense for a reversible process. In the limit where we apply the map an infinite amount of times, any value of q will be brought infinitely close to 0, which also does not sound like a reversible process.

[10] One should not confuse evolved/transported state variable with active/passive transformations or with Schroedinger/Heisenberg pictures. In those cases, the choice is between changing the state or the coordinates/observables. In our case, the state is always changing. The choice is between tracking the change with different values of the same state variable or different state variables that give the same value.

The issue is that while deterministic evolution will map states to states, not all bijective state to state maps can be considered deterministic and reversible.[11] This is a crucial point where this work is different from more usual treatments of dynamical systems. If one focuses only on the states themselves and counts them, it reaches the conclusion that a range of 1m contains the same number of initial conditions of 1Km: they have the same infinite number of elements. This conclusion has a number of problems: it does not match our physical intuition; it's not consistent with all types of topology; it does not explain why the laws of evolution are very often, not only continuous, but differentiable. In short: a bijection does not capture the entirety of the physical description.

If we work with a discrete topology then, by all means, the way to measure the size of a set is to count the elements in a set. But if we use the standard topology on $\mathbb{R}$, isolated points have measure zero (i.e. we can't physically distinguish them): what defines the number of initial conditions is the range of the state variables. As we'll see, the geometrical structures in Hamiltonian/Lagrangian classical/quantum mechanics are there to allow us to properly count states and, more specifically, possibilities along independent state variables. If we think that counting states is equivalent to counting elements in the set, we miss the whole point of that geometrical structure.

What makes the matter even more confusing is that the single physical idea of counting states and possibilities along independent state variables does not map to a single mathematical concept. Sometimes we need a measure, sometimes we need a symplectic form and sometimes we need a metric tensor. It will depend, case by case, on the state space of the system under study. But, however defined mathematically, deterministic and reversible motion must preserve that count. This is the crucial insight that gives physical meaning to those structures, that justifies the use of differential equations for state evolution, that tells us why symplectic and Riemannian geometry are so important in physics. Properly characterizing this count will be one of the key goals of the following sections.

V. COMPOSITE SYSTEMS AND REDUCIBILITY

Given that scientific reductionism (i.e. the idea of reducing physical systems and interactions to the sum of their constituent parts in order to make them easier to study) is at the heart of fundamental physics,(ref?) we now explore how to characterize a system in terms of its components. That is, we want to study the relationship between the state space of a composite system and the state space of its parts. In general, this is quite a complicated thing to do, which requires intimate knowledge of the system at hand. So we simplify our problem and study a material made of infinitesimal homogeneous parts. What we'll find is that under the additional assumption that the material is infinitesimally reducible (i.e. its state is equivalent to the states of its infinitesimal parts) and that each part undergoes deterministic and reversible evolution, the motion is suitably described by the standard framework of classical Hamiltonian mechanics.

V.A. Homogeneous decomposable systems and vector spaces

The notion that a system is decomposable means the states are equipped with a rule of composition that allows one to write $c = c_1 + c_2$: the composite system is the sum of its components. For example, the state of a ball is equal to the state of its top and bottom parts.

The notion that the system is homogeneous means that the states of the composite and of each part are not unrelated: they are all made of the same material.[12] In fact, the state space $\mathcal{C}$ of all systems composed of such material will include the state of the system as well as the states of its parts (i.e. $c, c_1, c_2 \in \mathcal{C}$). Since combining any two systems made of a homogeneous material will always give us a system made of the same homogeneous material (e.g. combining elements made of water gives us another element made of water), the state space $\mathcal{C}$ is closed under composition.

As we study the system under evolution, we'll also want to study state changes. For example, if a stable mixture of gas expands, we'll want to know the difference between the initial and final distributions. That is: $\delta c = \hat{c} - c$. Note that such a state difference may not describe a physical state: it may remove some material from one location to add it somewhere else. Yet, these state changes are still physically distinguishable objects that provide a configuration for the same homogeneous material, therefore we extend the state space $\mathcal{C}$ to include state differences.[13] All combined, this gives us the structure of an abelian group.

[11] In fact, we have already seen that the map must at least be continuous.

[12] Whether a system is homogeneous depends on context (e.g. air can be thought as homogeneous if the mixture of gases does not change in space or in time due to phase transitions or chemical processes) and one must check that such a property is maintained by time evolution.

[13] Note that one can characterize a state difference without knowing the states themselves. If we take some amount of material out of a container we know how the state changed, yet we may not know the amount of material before or after. In certain cases one may be interested in measuring the asymmetry of a particular state without measuring (or without being able to measure) the full state itself.

Proposition V.1. The state space $\mathcal{C}$ for a decomposable homogeneous material is an additive abelian (i.e. commutative) group.

Justification. We claim $\mathcal{C}$ is an additive monoid. There exists a law of composition $+ : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ that takes two states and returns one that is the physical composition of the two. The domain and codomain match because the material is homogeneous. The law is commutative $c_1+c_2 = c_2+c_1$ and associative $(c_1+c_2)+c_3 = c_1+(c_2+c_3)$, as it does not matter in what order we physically compose the components. There exists a unique zero element $c + 0 = c$ and it represents the physically empty state (i.e. no amount of material). $\mathcal{C}$ is an additive monoid by definition.

We claim $\mathcal{C}$ is an additive group. We require $\mathcal{C}$ to include state changes. A change of a physically distinguishable object is itself physically distinguishable therefore $\mathcal{C}$ is still a Hausdorff topological manifold as per IV.2. There exists an inverse $- : \mathcal{C} \to \mathcal{C}$ such that $c + (-c) = 0 \forall c \in \mathcal{C}$. $\mathcal{C}$ is an additive group by definition. □

V.B. Classical material and distributions

We also want to be able to express the state of the composite system in terms of the state of each part. For example, given the state of a fluid we'll want to know how the material is distributed in space. This in general will depend on how much the state of the composite system "knows" about its parts. For example, the position and orientation of an ideal rigid body is enough to define where all its constituents are, while the volume/pressure/temperature of an ideal gas is not enough to determine the position and momentum of all its constituents. Therefore we need to characterize the system further.

We will call a *particle* the smallest amount of a material to which we can assign an independent state. For example, a photon is a particle of light (it is the smallest amount of light we can describe), an infinitesimal amount of water is treated classically as a point particle (the smallest amount for a continuous fluid).[14]

We will call a *classical material* one that is homogeneous and infinitesimally reducible. That is, we can keep decomposing the system indefinitely into smaller and smaller parts, each with a well defined state. In this case, the particles are the infinitesimal parts given by the limit of this process of recursive reduction. Giving the state of the whole system is equivalent to giving the states of all particles. Given the state space $\mathcal{S}$ for the infinitesimal parts, which we assume consistently with IV.7 to be a manifold, a composite state $c \in \mathcal{C}$ will tell us the amount of material for each possible particle state. That is, each state is fully identified by a function $\rho_c : \mathcal{S} \to \mathbb{R}$ that returns the amount of material in the composite state c that is prepared in the particle state $\mathscr{s}$. For example, given a certain configuration of gas, we can tell the density of material that is at a specific point in space with a specific value of momentum.

The notion that the parts are infinitesimal requires the co-domain of ρ_c to be a real number, as opposed to an integer. The state space $\mathcal{S}$ could, instead, be a discrete set. For example, for a system composed of different tanks connected by pipes, the state of the composite system could be the overall distribution of water among the tanks. The amount in each tank is continuous (as we assume the water to be infinitesimally divisible) yet there are only a finite number of tanks the water can be placed into. As it links two physically distinguishable objects, ρ_c is a continuous function as discussed in IV.4.

As we combine the states of different parts, we sum the distributions over particle states $\rho_c(\mathscr{s}) = \rho_{c_1}(\mathscr{s}) + \rho_{c_2}(\mathscr{s})$: the amount of material in the composite state is the sum of the amount of material of the parts. We can also increase or decrease the amount of material by a constant factor $\rho_{c_1}(\mathscr{s}) = a\rho_{c_2}(\mathscr{s})$. This gives us the structure of a real vector space that is isomorphic to a subspace of continuous functions.

Proposition V.2. The state space $\mathcal{C}$ for a classical (i.e. homogeneous infinitesimally reducible) material is a vector space over $\mathbb{R}$ isomorphic to a subspace of the space of continuous functions. That is $\mathcal{C} \cong G \subseteq C(\mathcal{S}) \equiv \{\rho : \mathcal{S} \to \mathbb{R} \mid \rho \text{ is continuous}\}$, where $\mathcal{S}$ is the state space of an infinitesimal amount of material.

Justification. We claim $\mathcal{C}$ is an abelian group. $\mathcal{C}$ is the state space for a decomposable homogeneous material and is therefore an abelian group by V.1.

We claim T, the set of transformations that increase or decrease the amount of material in the system by a constant factor, is a field[15] isomorphic to $\mathbb{R}$. Consider $\tau : \mathbb{R} \to T$ the mapping between a number and the transformation that increases or decreases the amount of material by that factor. This transformation exists: the system is infinitesimally decomposable and the amount can be changed continuously. Define on T an addition $+ : T \times T \to T$ and a multiplication $* : T \times T \to T$ such that $\tau(a) + \tau(b) = \tau(a + b)$ and $\tau(a) * \tau(b) = \tau(a * b)$, $a, b \in \mathbb{R}$, so that the sum and product of the transformation is equal to the sum and product of their respective factors. τ is an isomorphism between T and $\mathbb{R}$ as fields.

We claim $\mathcal{C}$ is a vector space over $\mathbb{R}$. The abelian group $\mathcal{C}$ can be extended with the operations defined by T, as

[14] We use this definition as it allows us to retroactively talk on somewhat equal grounds about classical particles and quantum particles, and underscore how physically (and mathematically) they are very similar yet very different objects.

[15] Here field is intended in the abstract algebraic sense (a nonzero commutative division ring) which has no relationship to the field in the physics or differential geometry sense (a physical quantity/tensor with a value for each point in space).

each element $\tau \in T$ is a map $\tau : \mathcal{C} \to \mathcal{C}$. The map has the following properties: $(\tau_1 + \tau_2)\, c = \tau_1 c + \tau_2 c \ \forall \tau_1, \tau_2 \in T$ and $c \in \mathcal{C}$, increasing the amount of material by the sum of two constant factors is the same as combining the separate increases, and $\tau\,(c_1 + c_2) = \tau c_1 + \tau c_2 \ \forall \tau \in T$ and $c_1, c_2 \in \mathcal{C}$, increasing the amount of the total system is the same as the combination of the increased parts. $\mathcal{C}$ is a module over T, which is a field and isomorphic to $\mathbb{R}$. $\mathcal{C}$ is (isomorphic to) a real vector space.

We claim $\mathcal{C}$ is isomorphic to a subspace of $C(\mathcal{S})$ as a vector space over $\mathbb{R}$. $\forall c \in \mathcal{C}\ \exists! \rho_c : \mathcal{S} \to \mathbb{R}$ returning the amount of material for each state $\mathscr{s} \in \mathcal{S}$. $\rho_c \in C(\mathcal{S})$ by IV.4. Let $\varrho : \mathcal{C} \to C(\mathcal{S})$ such that $\varrho(c) \mapsto \rho_c$. As the system is reducible, two distinct composite states must represent different distributions. $\forall c_1, c_2 \in \mathcal{C}, c_1 \neq c_2 \implies \varrho(c_1) \neq \varrho(c_2)$. ϱ is injective. ϱ is a bijection between $\mathcal{C}$ and $\varrho(\mathcal{C})$. Let $c = c_1 + c_2$, then $\varrho(c) = \varrho(c_1) + \varrho(c_2)$ as the amount of material of the composition of two states is the sum of the individual amounts. Let $c_1 = \tau(a) c_2$, then $\varrho(c_1) = a\varrho(c_2)$ as $\tau(a)$ increases the amount of material by the factor a. ϱ is a homomorphism as a vector space. $\mathcal{C}$ is isomorphic to $\varrho(\mathcal{C}) \subseteq C(\mathcal{S})$. □

V.C. Integration and measure

When the topology of $\mathcal{S}$ is not discrete, the distribution ρ_c is a density, which is not something we directly measure. What we do measure experimentally are finite amounts of material. For example, the amount of fluid in a particular volume within a particular range of momentum specified in some units (e.g. moles, kg, ...). This means that given a composite state c and a set $U \subseteq \mathcal{S}$ of particle states compatible with an outcome of a process, we must be able to tell the amount of material that we will find associated with that outcome. That is for each open set $U \subseteq \mathcal{S}$ there will be a functional $\Lambda_U : \mathcal{C} \to \mathbb{R}$.

As we combine parts, or increase the amount of material in each part, the total amount of material found will have to be consistent with those operations. That is: $\Lambda_U(a_1 c_1 + a_2 c_2) = a_1 \Lambda_U(c_1) + a_2 \Lambda_U(c_2)$. For a proper state, one that is not the difference of two states, we will expect a positive amount of material. So, if a density ρ_c is positive everywhere, then the total amount is also positive. This makes Λ_U a positive linear functional. As the amount of material we measure is always finite, the functional applied to any composite state c over any set U will be finite.

The ability to associate finite amounts of material with sets of states allows us to define finite regions of $\mathcal{S}$ and compare them. Consider a region of position and momentum in phase space. If we are able to spread a finite amount of material into a non-infinitesimal uniform distribution, then we know we have a finite region. And the density will give us an indication of how big the region is: if we double the region, the density will halve. In other words: because we are describing densities over $\mathcal{S}$, we are able to give a unique measure[16] μ that allows us to count the number of states in each set of $\mathcal{S}$. In terms of said measure, positive linear functionals become integrals, $\Lambda_U(c) = \int_U \rho_c d\mu$.

This intuitive picture is formalized mathematically by the Riesz representation theorem for linear functionals(ref?), which gives $\mathcal{S}$ the structure of a measure space, with a suitable Borel σ-algebra and measure μ.

Proposition V.3. The state space $\mathcal{S}$ for the particles of a classical material is endowed with a Borel measure μ. The state space $\mathcal{C}$ for a classical material is isomorphic to a subspace of the space of Lebesgue integrable functions. That is $\mathcal{C} \cong G \subseteq L^1(\mathcal{S}, \mu) = \{\rho : \mathcal{S} \to \mathbb{R} \mid \int_{\mathcal{S}} |\rho| d\mu < \infty\}$.

Justification. We claim there exists a positive linear functional $\Lambda_U : \mathcal{C} \to \mathbb{R}$ for each $U \subseteq \mathcal{S}$ such that $\Lambda_U = \Lambda_{\mathrm{int}(U)}$ where $\mathrm{int}(U)$ is the interior of U. Let $\Lambda_U : \mathcal{C} \to \mathbb{R}$ be the functional that returns the amount of material in c compatible with the outcome associated with the open set $U \subseteq \mathcal{S}$. Λ_U is well-defined: U is associated with a physically distinguishable outcome. Λ_U is linear: $\Lambda_U(a_1 c_1 + a_2 c_2) = a_1 \Lambda_U(c_1) + a_2 \Lambda_U(c_2)$ as it has to be consistent with the operations of composing states and increasing/decreasing the amount of material by a factor. Λ_U is positive: if the value of the distribution for each particle state is positive then the total amount of material is positive. Let $U \subseteq \mathcal{S}$ not necessarily open. Define Λ_U as $\Lambda_{\mathrm{int}(U)}$.

We claim that $|\Lambda_U(c)| < \infty\ \forall c \in \mathcal{C}\ \forall U \subseteq \mathcal{S}$. Let $U \subseteq \mathcal{S}$ be an open set of particle states associated with an outcome. Let $c \in \mathcal{C}$ be a composite state. The amount of material of c associated with the outcome U must be finite, as physically we always work with finite quantities. $|\Lambda_U(c)| < \infty$. Let $U \subseteq \mathcal{S}$ not necessarily open. $|\Lambda_U(c)| = |\Lambda_{\mathrm{int}(U)}(c)| < \infty$.

We claim that $\Lambda_{U_1} + \Lambda_{U_2} = \Lambda_{U_1 \cup U_2} + \Lambda_{U_1 \cap U_2}\ \forall U_1, U_2 \in \mathcal{S}$. Let $U_1, U_2 \in \mathcal{S}$ be open sets of particle states associated with two outcomes. Suppose $U_1 \cap U_2 = 0$, $\Lambda_{U_1 \cup U_2} = \Lambda_{U_1} + \Lambda_{U_2}$ as the material found in $U_1 \cup U_2$ must be either in U_1 or U_2. Suppose $U_1 \cap U_2 \neq 0$, $\Lambda_{U_1 \cup U_2} = \Lambda_{U_1} + \Lambda_{U_2} - \Lambda_{U_1 \cap U_2}$ as the sum of the material associated with each outcome will double count the intersection. Let $U_1, U_2 \in \mathcal{S}$ not necessarily open. $\Lambda_{U_1} + \Lambda_{U_2} = \Lambda_{\mathrm{int}(U_1)} + \Lambda_{\mathrm{int}(U_2)} = \Lambda_{\mathrm{int}(U_1) \cup \mathrm{int}(U_2)} + \Lambda_{\mathrm{int}(U_1) \cap \mathrm{int}(U_2)} = \Lambda_{U_1 \cup U_2} + \Lambda_{U_1 \cap U_2}$

We claim that $\mathcal{S}$ is endowed with a unique Borel measure μ such that $\Lambda_U(c) = \int_U \rho_c d\mu$. $\mathcal{S}$ is locally compact as it is a manifold. $\mathcal{C} \cong G \subseteq C(\mathcal{S})$ therefore $\Lambda_U(c) \cong \Lambda_U(\rho_c)$. $\Lambda = \{\Lambda_U : C(\mathcal{S}) \to \mathbb{R}\}_{U \subseteq \mathcal{S}}$ is a family of positive linear functionals such that $\forall U \subseteq \mathcal{S}\ \Lambda_U = \Lambda_{\mathrm{int}(U)}$ and $\Lambda_{U_1} + \Lambda_{U_2} = \Lambda_{U_1 \cup U_2} + \Lambda_{U_1 \cap U_2}\ \forall U_1, U_2 \subseteq \mathcal{S}$. By

[16] Here measure is intended in the mathematical sense (a real valued function of a σ-algebra) which is distinct from other connotations in physics.

the extension A.1 of the Riesz representation theorem for linear functionals there exists a unique Borel measure μ such that $\Lambda_U(c) = \int_U \rho_c d\mu$.

We claim $\mathcal{C}$ is isomorphic to a subspace of $L^1(\mathcal{S}, \mu)$ as a vector space over $\mathbb{R}$. $\forall c \in \mathcal{C} \exists! \rho_c : \mathcal{S} \to \mathbb{R}$ returning the amount of material for each state $\mathscr{s} \in \mathcal{S}$. As $|\Lambda_U(c)| < \infty \ \forall c \in \mathcal{C} \ \forall U \subseteq \mathcal{S}$, then $\rho_c \in \mathcal{L}(S, \mu) \equiv \{\rho : S \to \mathbb{R} \mid \ |\int_U \rho d\mu| < \infty \ \forall U \subseteq S\}$. $\mathcal{L}(S, \mu) = L^1(\mathcal{S}, \mu)$ by A.2. Let $\varrho : \mathcal{C} \to L^1(\mathcal{S}, \mu)$ such that $\varrho(c) \mapsto \rho_c$. ϱ is a homomorphism, as justified in V.2. $\mathcal{C}$ is isomorphic to $\varrho(\mathcal{C}) \subseteq L^1(\mathcal{S}, \mu)$ as a vector space. □

V.D. Invariant densities and differentiability

On a state space with countable elements, using a single state variable, we have $\Lambda_U(c) = \sum_{q=a}^{b} \rho_c(q)$. We would expect the expression to become $\Lambda_U(c) = \int_a^b \rho_c(q) dq$ for states identified by a continuous state variable. By changing state variable, though, we see that this does not work in general: $\rho_c(\hat{q}) = \rho_c(q) dq/d\hat{q}$. This makes ρ_c a function not just of the state $\mathscr{s}$, but also of the state variables we are using. This is inconsistent with our previous definition. Moreover if the transformation is not differentiable, the density is not even well-defined. This tells us that $\mathcal{S}$ cannot be any manifold: it has to be one that allows state-variable-independent densities.

The first order of business is to guarantee that the density remains defined under an arbitrary change of state variables. This means the Jacobian of the transformation must exist. For the Jacobian to exist the transformation between state variables has to be differentiable. Therefore a manifold that guarantees densities to be well defined is one that guarantees that state variable transformations are differentiable: a differentiable manifold. $\mathcal{S}$ has a differential structure in the sense that the distributions and the volume element $d\mu$ are only defined on a set of state variables that are linked by differentiable transformations. In the same way that discontinuous changes of state variable can be detected because they tamper with physical distinguishability (i.e. the topology), non differentiable changes of state variable can be detected because they tamper with our ability to define state-variable-independent densities and count states.

Moreover, the density itself has to be differentiable. The distribution ρ_c is a real valued function of the state. As such, at least locally, we can use it as a state variable $q^1 = \rho_c$: we can identify particle states by the density of the material associated with them. This is more physically meaningful than one may first suspect, since placing physical markers (i.e. placing material at particular states) is a common way to define the references for a coordinate system. In a coordinate system where $q^1 = \rho_c$, the function is clearly differentiable as $\partial_1 \rho_c = 1$ and $\partial_i \rho_c = 0$ for $i \neq 1$. As we change state variables, the transformation is differentiable, therefore all the derivatives $\partial_{i'} q^1 = \partial_{i'} \rho_c$ exist. The density itself is differentiable.

Proposition V.4. The state space $\mathcal{S}$ for the particles of a classical material is a differentiable manifold. The state space $\mathcal{C}$ for a classical material is isomorphic to the space of Lebesgue integrable differentiable functions. That is $\mathcal{C} \cong C^1(\mathcal{S}) \cap L^1(\mathcal{S}, \mu)$, where $C^1(\mathcal{S}) \equiv \{\rho : \mathcal{S} \to \mathbb{R} \mid \rho$ is differentiable$\}$.

Justification. We claim $\mathcal{S}$ is a differentiable manifold. Let $\rho_c : \mathcal{S} \to \mathbb{R}$ be the distribution associated with a composite state $c \in \mathcal{C}$. Let q^i and $\hat{q}^j$ be two sets of independent state variables that fully identify states within $U \subseteq \mathcal{S}$. Let $\rho_c(q^i)$ and $\rho_c(\hat{q}^j)$ be the expression in local coordinates of the distribution. $\rho_c(\hat{q}^j) = \rho_c(q^i)|\partial_j q^i|$ as ρ_c transforms as a density. $\rho_c(\mathscr{s}) = \rho_c(q^i(\mathscr{s})) = \rho_c(\hat{q}^j(\mathscr{s}))$. The Jacobian $|dq'/dq|$ exists and is non-zero. The map between any two charts is differentiable. $\mathcal{S}$ is a differentiable manifold.

We claim that the distributions associated with composite states are differentiable in the regions where they are strictly monotonic. Let ρ_c be the distribution associated with state c. Let ρ_c be strictly monotonic in a region $U \subseteq \mathcal{S}$. Then its fibers in U, the inverse images of the values, are connected hypersurfaces. We can construct a chart such that $q^1 = \rho_c$. The distribution is differentiable over U for that variable: $\partial_i \rho_c = \partial_i q^1 = \delta_i^1$ where δ_i^j is the Kronecker delta. Let q^j be another set of state variables. $q^1(q^j)$ is differentiable as the manifold is differentiable. $\rho_c(q^j) = q^1(q^j)$ is differentiable over U.

We claim that the distributions associated with composite states are differentiable. Let ρ_c be the distribution associated with state c. Let ρ_{c_0} be the distribution associated with state c_0. Let ρ_{c_0} be strictly monotonic in a region $U \subseteq \mathcal{S}$. Let $\delta\rho = \rho_{c_0} - k\rho_c$ where $k \in \mathbb{R}$ and $k > 0$. Let k be sufficiently small so that $\delta\rho$ is strictly monotonic over U. ρ_{c_0} and $\delta\rho$ are differentiable over U as they are strictly monotonic over U. ρ_c is a linear combination of differentiable functions over any arbitrary region U. ρ_c is differentiable.

We claim $\mathcal{C} \cong C^1(\mathcal{S}) \cap L^1(\mathcal{S}, \mu)$. Let ρ_c be the distribution associated with a state c. Let $\rho \in C(\mathcal{S}) \cap L^1(\mathcal{S}, \mu)$ be a Lebesgue integrable continuous function. If $\rho_c \neq \rho$ then $\exists U \in \mathcal{S}$ such that $\int_U \rho_c d\mu \neq \int_U \rho d\mu$. Both integrals are finite as $\rho_c, \rho \in L^1(\mathcal{S}, \mu)$. Both integrals are expressible with the same state variables as all elements are defined on the same differential structure. Therefore there exists an outcome for $\mathcal{S}$ that can physically distinguish the two distributions. There must be a state $c_1 \in \mathcal{C} \mid \varrho(c_1) = \rho$. $\varrho : \mathcal{C} \to C(\mathcal{S}) \cap L^1(\mathcal{S}, \mu)$ is surjective. ϱ is a homomorphism, as justified in V.2. ϱ is an isomorphism between $\mathcal{C}$ and $C(\mathcal{S}) \cap L^1(\mathcal{S}, \mu)$ as vector spaces. □

As we identified the space of Lebesgue integrable differentiable functions $C^1(\mathcal{S}) \cap L^1(\mathcal{S}, \mu)$ as the space of distributions that are physically meaningful, we can better understand why other commonly used function spaces do not fit the bill. Some are not restrictive enough. The set of Lebesgue integrable functions $L^1(\mathcal{S})$ includes dis-

continuous functions that for IV.4 are unphysical. The space of continuous functions $C(\mathcal{S})$, the space of continuous functions that vanish at infinity $C_0(\mathcal{S})$ and the space of differentiable functions $C^1(\mathcal{S})$ include functions whose integral is infinite, which would represent an infinite amount of material. Some definitions are too restrictive. Requiring compact support (i.e. the function is different from zero only on a finite region) would exclude distributions, such as Gaussians, that span over the whole range of states. Such distributions are physical: if we take a finite volume of an ideal gas at equilibrium, the momentum distribution spans over the whole range. Schwartz space $S(\mathcal{S})$ excludes functions that are not infinitely smooth which we don't have a general physical justification for.[17]

Another consideration is that while the norm associated to $L^1(\mathcal{S})$ is $\int_{\mathcal{S}} |\rho| d\mu$, the expression $\int_{\mathcal{S}} \rho d\mu$ may be more physically meaningful in some cases. For proper states, the two are the same and they represent the total amount of material. For state changes they differ. The second represents the amount of material added (or taken away if negative) by the state change. For example, if $\delta\rho$ is the change due to deterministic and reversible evolution, the amount of material does not change and therefore $\int_{\mathcal{S}} \delta\rho \, d\mu = 0$. The first norm would represent the total amount of material that is changing (i.e. being added and being removed), therefore $\int_{\mathcal{S}} |\delta\rho| \, d\mu = 0$ means no material is moving, no change is occurring.

Note that neither of those expressions matches the norm given by the inner product defined as $\langle f, g\rangle = \int_{\mathcal{S}} fg \, d\mu$. Moreover $\mathcal{C}$, since it excludes discontinuous functions, can never be a complete normed space: it will not include the limit for all Cauchy sequences; it cannot be a Banach or a Hilbert space. Such a construction, though, is still useful. Consider the following expressions:

$$\Lambda_{\mathcal{S}}(c) = \int_{\mathcal{S}} \rho_c d\mu = \langle 1, \rho_c\rangle$$

$$\Lambda_U(c) = \int_U \rho_c d\mu = \langle 1_U, \rho_c\rangle$$

$$\rho_c(q_0, p_0) = \int_{\mathcal{S}} \delta_{q_0,p_0} \cdot \rho_c d\mu = \langle \delta_{q_0,p_0}, \rho_c\rangle$$

$$p_{total}(c) = \int_{\mathcal{S}} p \cdot \rho_c d\mu = \langle p, \rho_c\rangle$$

$$q_{avg}(c) = \frac{\int_{\mathcal{S}} q \cdot \rho_c d\mu}{\int_{\mathcal{S}} \rho_c d\mu} = \frac{\langle q, \rho_c\rangle}{\langle 1, \rho_c\rangle}$$

For each state, they represent respectively the total amount of material, the material within U, the density at (q_0, p_0), the total momentum and the average of a state variable. As they are linear functionals of $\mathcal{C}$, they can be expressed using their dual vector. But these vectors $(1_U,\ p,\ ...)$ are *not* elements of $\mathcal{C}$. They do not represent states, they are not continuous and integrable, and there is no single general physical meaning for all of them. Therefore, as we mentioned before, while we can extend the function space for convenience, computation purposes or to study a limit case, we need to be mindful of the extension and carefully check that the mathematical results are physically meaningful case by case.

V.E. Invariant densities and cotangent bundle

Now that we are guaranteed that the density is properly defined for all state variables, we have to make sure it is invariant. The distribution should only depend on the states, and not the particular choice of state variables to label them.

If ρ_c is to remain invariant under state variable changes, not only does the Jacobian have to exist but it must be unitary. This means we cannot simply change one state variable as we please. If we change one at least another has to change in some coordinated way such that the Jacobian of the total transformation is unitary. This means that we cannot change physical units of all the variables as we like: they are part of a unit system.

The simplest case is when a single variable is enough to define our units, and therefore tell how all the others must change. Suppose that our state space $\mathcal{S}$ is identified by n independent state variables $\{q, k^i\}$ and a change of units for the first, that is $\hat{q} = \hat{q}(q)$, determines the change for all others. The Jacobian matrix is the block matrix:

$$J = \begin{bmatrix} d_q\hat{q} & 0 \\ \partial_q\hat{k}^j & \partial_i\hat{k}^j \end{bmatrix}$$

The Jacobian has to be unitary, therefore all elements $\partial_i\hat{k}^j$ must be well defined and $|\partial_i\hat{k}^j| = 1/d_q\hat{q}$. As this can only happen if $\partial_i\hat{k}^j$ is a 1×1 matrix, there can only be one k. This gives $\partial_k\hat{k} = d_{\hat{q}}q$ and $\hat{k} = d_{\hat{q}}q\, k + f(q)$. As this is a change of units, we expect the zero value of k to remain constant: $\hat{k}(q, 0) = 0 = f(q)$. Therefore $\hat{k} = d_{\hat{q}}q\, k$ transforms as the component of a covariant vector. $\mathcal{S}$ is isomorphic to $T^*\mathcal{Q}$, the space of co-vectors at a point, where $\mathcal{Q}$ is the manifold that defines our unit.

Physically, this means that the state variable k, which is the classical analogue of the wave number, is expressed with the inverse of the unit used for q. If q is in meters, k is in inverse meters. Consider now the area $dq \wedge dk$ of an infinitesimal rectangular region. This quantity is dimensionless and therefore invariant. The total number of states will be proportional to it, as doubling the range of dq or dk will give us double the number of possible states. We have $d\mu = \hbar dq \wedge dk$, where $\hbar$ is the proportionality constant that will depend on the unit chosen to count the possibilities of the pair (q, k). That is, a unitary range of

[17] One can, though, make the argument that given any distribution ρ one can find an infinitely smooth ρ_{sm} such that the difference in description is small compared to the error already introduced by assuming the system to be homogeneous and infinitely reducible. The Whitney approximation theorem(ref?) makes this mathematically well defined.

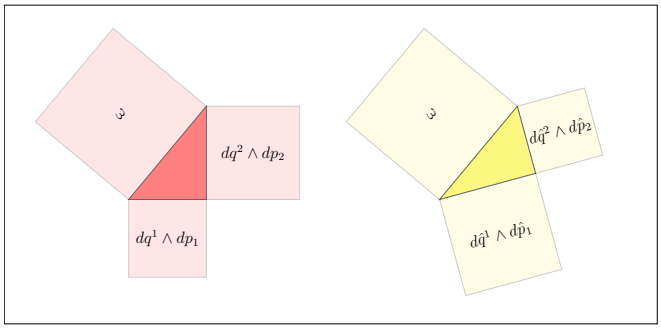

FIG. 2. The symplectic form ω and its components along two independent degrees of freedom. As we change state variables, the new d.o.f. are also independent and therefore orthogonal to each other. Note: this is a 2D conceptualization of 4D space.

possibilities for q and k will give us $\hbar$ possibilities.[18] The state count in a finite (i.e. compact) region $U \subseteq \mathcal{S}$ will be given by $\mu(U) = \int_U \hbar dq \wedge dk$: we are finally able to write the volume of integration in terms of state variables.

We can generalize to the case where $\mathcal{Q}$ has more than one dimension. As we must be able to change one unit at a time, independently from the other, to each q^i will correspond a k_i that uses the inverse of the corresponding units. We have $\mathcal{S} \cong T^*\mathcal{Q}$ and $d\mu = \hbar^n dq^n \wedge dk_n$. But it's not just the volume, the state count, that is preserved as we change state variables. Each pair (q^i, k_i) is expressed by an independent unit with the count of possibilities given by $\hbar dq^i \wedge dk_i$. We call such a pair a degree of freedom. As they are independent the total state count is the product of the possibilities of each d.o.f.: $d\mu = \hbar^n dq^n \wedge dk_n = \prod_{i=1}^{n} \hbar dq^i \wedge dk_i$. In other words: independent d.o.f. are orthogonal surfaces in $\mathcal{S}$. The possibility count of an arbitrary degree of freedom, then, will be $\sum_{i=1}^{n} \hbar dq^i \wedge dk_i$, the sum of the projections over the n orthogonal and independent d.o.f. defined by (q^i, k_i).

The possibility count for each d.o.f. (i.e. the wedge product within a d.o.f.) and the orthogonality of different d.o.f. (i.e. the scalar product across d.o.f.) must be the same regardless of the choice of state variables. We can express both requirements mathematically in a compact way. We first define conjugate momentum as $p_i = \hbar k_i$ and unified state variables as $\xi^a \equiv \{q^i, p_i\}$. Then we consider the canonical symplectic two-form $\omega = \Sigma\, dq^i \wedge dp_i$ given by the following components:

$$\omega_{ab} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \otimes I_n = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$$

It returns the wedge product within a d.o.f. and the scalar product across. Requiring the invariance of this metric under state variable changes assures us the count of states and possibilities is well defined.

This gives us insight on the physical meaning of the geometrical structure $T^*\mathcal{Q}$. The canonical one-form $\theta = kdq$ represents the geometrical object we associate with each particle state. The symplectic form ω is the bilinear that quantifies the possibilities described by two given state variables. Note, instead, that the relationship $\omega = -\hbar d\theta$, while mathematically true, has no clear physical meaning as we have not defined what the exterior derivative on a state actually represents.

We can capture the above discussion by stating that the state space $\mathcal{S}$ is the symplectic manifold $(T^*\mathcal{Q}, \omega)$, where $\mathcal{Q}$ is the manifold that defines the unit system. This is the only manifold that allows us to define state-variable-invariant densities. The symplectic form allows us to count possibilities on an arbitrary d.o.f.

Proposition V.5. The state space $\mathcal{S}$ for the particles of a classical material is a symplectic manifold formed by a cotangent bundle $T^*\mathcal{Q}$ equipped with the canonical two-form ω.

Justification. We claim the simplest (i.e. lowest dimension) state space $\mathcal{S}$ that allows state-variable-invariant densities is the cotangent bundle $T^*\mathcal{Q}$ of a single dimensional manifold $\mathcal{Q}$. Consider a set of state variables for $U \in \mathcal{S}$: the simplest unit system is one defined by a single state variable. Let q be the state variable that defines the unit system. Let $\mathcal{Q}$ be the manifold charted by q. A diffeomorphism $\hat{q} = \hat{q}(q)$ fully defines a change of units. Let (q, k^i) with $i = 1...n-1$ be a set of state variables. The Jacobian determinant $|J| = d_q\hat{q}|\partial_i \hat{k}^j| = 1$ as densities must be invariant. $|\partial_i \hat{k}^j| = d_{\hat{q}}q$. $\partial_i \hat{k}^j$ is one dimensional as its components must be fully specified by the previous relationship. Let $k = k^1$. $\partial_k \hat{k} = d_{\hat{q}}q$. $\hat{k}(q, 0) = 0$ as a change in units does not change the 0 value. $\hat{k} = d_{\hat{q}}qk$. k changes as the component of a co-vector. Each state is identified by a point and a co-vector in $\mathcal{Q}$. The simplest state space $\mathcal{S}$ is isomorphic to $T^*\mathcal{Q}$ where Q is the one dimensional manifold that defines the units.

We claim the simplest state space $\mathcal{S}$ that allows state-variable-invariant densities is the symplectic manifold $(T^*\mathcal{Q}, \omega)$ where $T^*\mathcal{Q}$ is the cotangent bundle of a one dimensional differentiable manifold $\mathcal{Q}$ and ω is the canonical two-form. Express integration in coordinates. $\int_{\mathcal{S}} \rho_c d\mu \propto \int_{\mathcal{S}} \rho_c(q, k) dq \wedge dk$. $d\mu = \hbar dq \wedge dk = dq \wedge dp = \omega$ where $\hbar$ is a constant, $p \equiv \hbar k$ and ω is the canonical two form. ω is invariant under state variable change. ω is the symplectic form for $\mathcal{S} = T^*\mathcal{Q}$. $\mathcal{S} = (T^*\mathcal{Q}, \omega)$ is a symplectic manifold.

[18] The actual value and physical dimensions of $\hbar$ are determined by the system of units, and should not be taken to describe some intrinsic physical property. Only dimensionless relationships to other physical constants, such as the fine structure constant $\alpha = e^2/\hbar c$, can possess that trait. That is why one can choose "natural units" for which $\hbar = 1$. In SI units the relationship $\hbar dk = dp = mdu$ between kinetic and conjugate momentum, derived later, sets the relationship between units.

We claim the state space $\mathcal{S}$ for the particles of a homogeneous classical material is a symplectic manifold $(T^*\mathcal{Q}, \omega)$ where $T^*\mathcal{Q}$ is the cotangent bundle of an n-dimensional differentiable manifold $\mathcal{Q}$ and ω is the canonical two-form. Let q^i be a set of n continuous independent state variables that define the units necessary to describe a state in $U \subseteq \mathcal{S}$. Let $\mathcal{Q}$ be the manifold charted by q^i. Locally $\mathcal{Q} \cong \prod \mathcal{Q}^i \cong \mathbb{R}^n$. Changing units of one state variable must not change the units of the other as they are independent. State-variable-distribution can be defined separately on each degree of freedom. For each $\mathcal{Q}^i$ we have a covector $k_i(q)dq^i \in T^*\mathcal{Q}^i$. Locally $\mathcal{S} \cong \prod T^*\mathcal{Q}^i \cong T^* \prod \mathcal{Q}^i \cong T^*\mathcal{Q}$. Integration must also be defined on an independent d.o.f. There must exist a non-degenerate two-form ω such that $\int_{U \subset \mathcal{S}} \rho_c \omega$ where U is any two dimensional surface in $\mathcal{S}$. ω has to be form invariant under state variable change. The canonical two-form is the only such form. $\omega = \sum dq^i \wedge dp_i$. $\mathcal{S} = (T^*\mathcal{Q}, \omega)$ is a symplectic manifold. □

It should be clear by now that the case of discrete topology is qualitatively different from the standard topology for $\mathbb{R}^n$. The notion that the continuous case is a limit of the discrete case, that it's "like the discrete but with more points", leads in this case to erroneous intuition. The key question is: can we physically distinguish an isolated state? Can we have an outcome associated with only one element? The answer is no with the standard topology on $\mathbb{R}^n$. The consequence is that when we define the measure μ for the count of continuous states, we assign a finite value to a finite region, and we assign zero measure to a single state. That is what gave us integration, densities and, ultimately, conjugate variables. If we were to use the discrete topology on $\mathbb{R}^n$, if we were to assume we can identify single states as we do for countable states, then we would assign measure one to each state and infinity to a finite region. We would not have integration, just a simple sum. Our distribution would not be a density and there would be no justification for conjugate variables. A finite distribution could only distribute a finite amount of material in a finite number of states.

In other words: this is not a case where a difference at small scales leads to a small difference at large scales. The two cases are radically different. We can tell them apart. We should take the use of densities and conjugate quantities as evidence that quantities like space and time are not discrete in the sense that the processes we use to distinguish those quantities cannot identify single instances. The quantum case does not change this, as the use of densities (in the form of the wave function) and conjugate quantities is even more prominent.[19]

[19] If one wants to posit that space is fundamentally discrete at scales where both classical and quantum mechanics cease to be valid, the better strategy would be to work with a discrete version of $T^*\mathcal{Q}$ and not $\mathcal{Q}$ by itself. Though it is not clear how a discrete topology would become non-discrete in the limit, at least the measure may be workable.

It should also be noted that classical particles, under this light, cannot be considered point-like. As they are the limit of infinitesimal subdivision, their spatial extent becomes infinitesimal but not zero.[20] But suppose particles were truly point-like. Then distinguishing particles would mean distinguishing points. We are back to the idea of a discrete topology on $\mathbb{R}^n$. In that case we would not have conjugate momentum, no $T^*\mathcal{Q}$, just the coordinates of the point in $\mathcal{Q}$. This would actually be more self-consistent: why wouldn't a point be enough to define the state of a point-like particle? As before, we should take the use of conjugate quantities as evidence that classical particles are really the limit of an infinitesimal subdivision.

Finally, we should note that we are in a position similar to the one discussed in IV.9 for discrete states. We have a topological space plus a measure that allows to count states and a symplectic form that allows to count possibilities, both expressible in terms of state variables. We recovered this structure starting from the idea of an infinitesimally reducible system. That led to a state space $\mathcal{S}$ for the infinitesimal parts, on which we must be able to define state-variable-independent distributions, which in turn gave us degrees of freedom made by pairs of conjugate variables and the symplectic form typical of classical phase space. In short: being able to measure the amount of material is what allows us to count states.

V.F. Infinitesimal reducibility

Now that we have fully characterized what we mean by a classical material, we can stipulate the following:

Assumption II (Infinitesimal reducibility)**.** *The system under study is composed of an infinitesimally reducible homogeneous material and each part undergoes deterministic and reversible evolution.*

Rationale. The idea is that time evolution specifies a map for the state space $\mathcal{S}$ of each infinitesimal part. Knowing how the parts evolve tells us how a composite state c evolves as well.

Consistent with what we said in assumption I, if we defined a state for each particle, then a deterministic and reversible evolution on that state must exist. Yet, the idea that we can assign states to infinitesimal parts should be considered only a *simplifying* assumption. The obvious reason is that we know this does not work in practice: as we keep decomposing the material we end up with molecules, atoms and then subatomic particles. But it is instructive to understand when and how the assumption breaks down at a more conceptual level.

[20] This picture is also compatible with general relativity, unlike point-like particles. These would not follow geodesics as their infinite mass density would significantly affect the gravitational field.

The first problem is methodological. To be able to talk about the states of a part we need a physical process that is able to distinguish between them. For a billiard ball we can imagine marking one spot with a red marker. This allows us to track it as the ball moves or collides with other balls. For an electron, instead, we do not have a way to mark a piece. In fact when two electrons scatter we can't even tell which is which, let alone what portion went where. The classical assumption may not hold because we do not have suitable physical processes at our disposal.

Even if we are able to track parts, the assumption requires them to be infinitesimal, the limit of a process of infinite recursive subdivision. The best we can do experimentally is to confirm that the assumption holds up to the smallest precision available. Therefore even in the best of cases it cannot be considered an experimentally validated assumption but a reasonable default assumption (e.g. "it worked so far").

The second, more conceptual, problem is that the idealizations that allowed us to define states in assumption I may not hold as the parts get smaller. Recall the cannonball whose motion is sufficiently unaffected by the photons that scatter off its surface. As we consider smaller and smaller parts, at some point we will find an amount of material that is affected by the interaction with the air or photons scattering off of it. At that point the parts are no longer sufficiently isolated to define an independent state, their evolution depends on the environment. Therefore the deterministic evolution of the whole cannot be reduced to the deterministic evolution of its parts.

Another issue is that defining a state requires some kind of asymmetry between system under study and environment. The future state of the system does not depend on the state of the environment, so that we can define an independent state, yet the future state of the environment is affected by the state of the system, so that we can have external processes that allow us to define physical distinguishability. If we try to assume that both the system and environment are really made of the same classical material, then the claim to that asymmetry is lost: the infinitesimal parts of both system and environment must affect one another in the same way.

Another issue arises if we include the measuring device in our description. Ideally, we'd like to require the following:

- The system under study is deterministic and reversible
- The measuring device ascertains the state of the system
- System and measuring device, as a whole, is deterministic and reversible

Unfortunately, these three requirements together are inconsistent. Say $a \in A$ is the initial state of our system and $b \in B$ the initial state of our measuring device: the final state $\hat{a} = \hat{a}(a)$ because the system is deterministic; $a = f(\hat{b})$ because we are able to know the initial state of the system by looking at the final state of the measuring device; $b = b(\hat{a}, \hat{b})$ because the system as a whole is reversible. Since $\hat{a}$ is determined by a, and a can be known from $\hat{b}$, $b = b(\hat{a}(f(\hat{b})), \hat{b}) = b(\hat{b})$ can be determined from just $\hat{b}$. But since $a = f(\hat{b})$, the whole past state could be reconstructed just by looking at the future state of the measuring device. But $(\hat{b}) \to (a, b)$ cannot be an injection, therefore we arrive at a contradiction.

Of those three requirements we can only pick two. If we want determinism and reversibility for the combined system, either both systems are not independently deterministic or the second is not a measuring device. In other words: we can't expect to have deterministic and reversible evolution at all levels of aggregation and have the parts interact in any physically meaningful way.

The same problem of physically meaningless interaction surfaces when we consider the states of infinitesimal parts. Under our classical assumption, if we were to take the state of an infinitesimal part to really be its full description, with no unstated part, then each piece would evolve independently, oblivious to the other parts. Each particle would essentially reside in its own separate physical universe, as it cannot physically distinguish anything else. This is not physically meaningful.

If we assumed the system is deterministic and reversible only as a whole, then each part could evolve depending on the states of the other parts. This would seem much better, as the state of the pieces is still exhaustive yet they are physically connected to each other. But this does not actually solve the problem of independent oblivious components. First, for this to work, one would have to specify how the states of the parts were defined since their evolution is individually no longer deterministic (the future of each part depends on the state of other parts). But ignoring that issue, the bigger problem is that we can always locally separate the evolution into independent degrees of freedom. For example, the position and momentum of two particles may affect each other during the evolution, but the average and difference in position and momentum may evolve independently.[21] We can always find such a local decomposition, and the easiest way to see that is in terms of the transported variables: they retain the original value; they clearly evolve independently. And since the pieces are infinitesimal, a local decomposition is all that is needed. While such description may be cumbersome to achieve in practice, conceptually it is still possible. So, even if each particle evolution depends on the state of the others, the system is described by degrees of freedom that evolve independently, oblivious to each other.

The moral is that the classical idea of being able to assign to all systems a state which represents their full description does not work. The state is only what we can

[21] Also note that any Hamiltonian is locally isomorphic to a free particle (ref?)

describe and it can't be the full description. As the division between system and environment is subjective, each system must be able to function as both. Therefore it will have a part whose evolution depends only on the system itself, the state, and a part whose evolution depends on other systems, the unstated part. While they may not be the same in all circumstances, they must exist. The state is what gives the identity to the system, the part we can study and describe. The unstated part is what allows continuous interaction between the system and the environment; it's what allows us to study and describe the state. As we'll see later, it is precisely this problem that quantum mechanics conceptually addresses better than classical mechanics.

One final problem is with time itself: an infinitesimally reducible system undergoing deterministic and reversible evolution cannot tell time. If it did, we would be able to find a quantity that changes through time but is invariant under state variable changes. The problem is that deterministic and reversible time evolution is equivalent to a state variable change. In fact, at a fixed time, we can choose to describe the system with the transported state variables of any past or future times. So, all quantities that are invariant under state variable changes are also invariant under time evolution. The somewhat ironic result is that while time is essential for the very definition of deterministic and reversible evolution, time itself has to be defined by some other type of process.[22]

In light of what we discussed, we cannot take the classical assumption to be fundamental, in the sense that we cannot take it to strictly apply to all of the universe. While ultimately flawed, the assumption can be considered valid for a great number of macroscopic systems, and is also useful as a default assumption of sorts. Understanding its shortcomings will help us later to see how the quantum assumption solves, at least partially, some of these issues.

As a final note: we caution against automatically thinking that the classical assumption is valid for all macroscopic systems. It is conceptually possible to have a macroscopic system where a clear independent state cannot be assigned to each part, in which case the assumption would not hold.(ref?) □

V.G. Hamiltonian mechanics

We are finally ready to write the equations of motion. As per IV.3 our evolution is at least a self-homeomorphism $f : \mathcal{S} \leftrightarrow \mathcal{S}$. But this is not enough.

The evolution must map the distribution point-wise so that the value associated at the initial state is the same as the value at the final state. All the material that starts in $\mathscr{s}$ has to end up in $\hat{\mathscr{s}}$. In math terms $\rho_{\hat{c}}(\hat{\mathscr{s}}) = \rho_c(\mathscr{s})$.

In the same way we expect the total amount of material to be conserved. If $U \in \mathcal{S}$ is a set of initial particle states and $\Lambda_U(c)$ is the amount of material associated with that set, then we expect it to be equal to the amount of material $\Lambda_{\hat{U}}(\hat{c})$ associated with the set of final states $\hat{U} \in \mathcal{S}$.

But probably the easiest way to look at it is that initial and final sets of states have to possess the same count of states and possibilities as defined by the symplectic form ω. Therefore the area within a degree of freedom, which represents the number of possibilities in said d.o.f., needs to be mapped to an equal area within the transported d.o.f. (i.e. the d.o.f. defined by the transported state variables). Independent d.o.f. must remain independent, and therefore transported orthogonal d.o.f. remain orthogonal. This means that the product of possibilities of independent d.o.f. is also conserved. These statements are the physical justification of Gromov's non-squeezing theorem [13–15] and Liouville's theorem. And they give intuitive insight on the geometry of Hamiltonian systems.

Mathematically, under the classical Hamiltonian assumption, deterministic and reversible evolution is a self-symplectomorphism (or self-isometry or canonical transformation depending on your math training). That is, it does not just preserve the topology but also the symplectic form ω.

Proposition V.6. A deterministic and reversible evolution map for the particles of a classical material is a self-symplectomorphism. That is: $\mathcal{T}_{\Delta t} : T^*\mathcal{Q} \to T^*\mathcal{Q}$ and $\mathcal{T}^*_{\Delta t}\omega = \omega$ where $\mathcal{T}^*_{\Delta t}$ is the pullback of $\mathcal{T}_{\Delta t}$.

Justification. We claim $\mathcal{T}_{\Delta t}$ is a self-homeomorphism on $T^*\mathcal{Q}$. The state space for the particles of a classical material is $T^*\mathcal{Q}$ by V.5. $\mathcal{T}_{\Delta t}$ is a deterministic and reversible evolution map and by IV.3 is a self-homeomorphism.

We claim $\mathcal{T}_{\Delta t}$ is a symplectomorphism on $(T^*\mathcal{Q}, \omega)$. The distribution on final states must be defined therefore the Jacobian for $\mathcal{T}_{\Delta t}$ exists and is non-zero: $\mathcal{T}_{\Delta t}$ is a diffeomorphism. A deterministic and reversible process conserves the number of states and possibilities. ω is the two-form that returns the count of possibilities. ω is invariant under deterministic and reversible evolution. $\mathcal{T}_{\Delta t}$ is a symplectomorphism by definition. □

If we assume continuous time evolution we have the following:

Proposition V.7. A continuous deterministic and reversible evolution for the particles of a classical material admits a Hamiltonian $H \in C^2(T^*\mathcal{Q}, \mathbb{R})$ that allows us to write the laws of evolution as

$$d_t q^i = \partial_{p_i} H$$
$$d_t p_i = -\partial_{q^i} H$$

Justification. We claim the vector field $S \in T\mathcal{S}$ for the infinitesimal displacement $S^a = d\xi^a/dt$ admits a potential $H \in C^2(T^*\mathcal{Q}, \mathbb{R})$ such that $S^a\omega_{ab} = \partial_b H$. The state

[22] For example, if the phase-space volume occupied by a distribution increased, it would give us a coordinate-invariant quantity that changes in time. This, however, is not possible under Hamiltonian evolution.

space $\mathcal{S}$ for the particles of a classical material is a symplectic manifold by V.5. The map for infinitesimal time evolution $\mathcal{T}_{dt}$ is an infinitesimal self-symplectomorphism by V.6. By A.3 the infinitesimal displacement S admits a potential $H \in C^2(T^*\mathcal{Q}, \mathbb{R})$ such that $S^a\omega_{ab} = \partial_b H$

We claim the state variables evolve according to Hamilton's equations. $S^a\omega_{ab} = d_t\xi^a\omega_{ab} = \partial_b H$. For $b = \{1, ..., n\}$ we have $d_t p_i(-1) = \partial_{q_i} H$. For $b = \{n + 1, ..., 2n\}$ we have $d_t q^i(+1) = \partial_{p_i} H$. □

We recognize the familiar set of Hamilton's equations. They are the set of equations that describe the deterministic and reversible motion of the infinitesimal parts of an infinitesimally reducible homogeneous material. Note that the argument goes the other way as well. A system governed by Hamilton's equations is deterministic and reversible: the equation of motion given by H are differentiable since H is twice differentiable. Therefore the equations are Lipschitz continuous and, by the Picard-Lindelöf theorem(ref?), a unique solution exists.

In other words: the forces that conserve energy (i.e. the value of a suitable Hamiltonian) are exactly the ones that provide deterministic and reversible motion. The challenge in their derivation mainly lies in the necessary use of different branches of mathematics (e.g. topology, measure theory, differential geometry, symplectic geometry). The conceptual meaning, on the other hand, is hopefully straightforward.

It's important to realize that, during the derivation, multiple mathematical features (e.g. invariant densities, cotangent bundle for phase space, symplectomorphism) were justified by the same physical assumption. The math itself gives us no indication that the different features stem from the same source; therefore, the math itself does not give us a conceptually unified picture. This is one of the reasons we believe that starting from the physical description is objectively better if we are to come to a better understanding of our physical theories.

Note that we could have taken different approaches. For example, we could have appealed to information theory and required our invariant distributions to preserve Shannon's information entropy[11], as no information should be gained or lost during a deterministic and reversible process. Or we could have appealed to statistical mechanics and required that the determinant of the covariance matrix be conserved, as a deterministic and reversible process should be defined at the same level of precision. With suitable treatment of independent d.o.f. both these approaches recover Hamiltonian mechanics as well. While the full treatment is outside the scope of this work, we want to underline that the definitions used in this work go a long way to building bridges at the core of different mathematical and scientific branches.

VI. TIME DEPENDENT EVOLUTION

We now generalize our discussion to include time dependent dynamics. This is needed when the evolution map is not the same at all instants or when state variables depend on time (e.g. changing to a moving frame).

To do this we will redefine the particle state space to include the temporal degree of freedom. We'll find that the dynamics is described by relativistic Hamiltonian mechanics in the extended phase space. We'll also find that particle states divide into standard and anti-states depending whether time is aligned or anti-aligned with the evolution parameter. Note that no extra assumption is needed, which makes relativistic mechanics simply the correct time dependent description of the deterministic and reversible evolution of an infinitesimally reducible system.

VI.A. Time changes

The discussion so far has been limited to the case where both the state variables and the evolution map never change throughout the evolution. This is too restrictive as there are very reasonable situations in which this does not hold.

The first obvious case is that the map may be time dependent. This does not affect our definition of determinism and reversibility as we still can tell final state from initial state and vice-versa. But at this point we have no way to specify a time dependent evolution: the Hamiltonian $H : T^*\mathcal{Q} \to \mathbb{R}$ as we derived it is just a function of the state.

The second case is when we perform a transformation for the time parameter $\hat{t} = \hat{t}(t)$. The equations of motion transform and we have $d_t\hat{t}\, d_{\hat{t}} q^i = \partial_{p_i} H$. $q^i : T^*\mathcal{Q} \to \mathbb{R}$ and $H : T^*\mathcal{Q} \to \mathbb{R}$ are real functions of the state space and cannot be redefined to include $d_t\hat{t}$. Physically the transformation is introducing fictitious forces that are not conservative, so they cannot be expressed by a Hamiltonian. But this means that we have an ill defined mathematical framework as the notion of determinism and reversibility is not defined in a way that is independent of time transformations. We need a framework that is capable of handling the fictitious forces as well.

The third case is when we perform a state variable transformation $\hat{q} = \hat{q}(q, t)$ that is time dependent (e.g. $\hat{q} = q + vt$). Our composite state distribution $\rho(q)$ becomes $\rho(\hat{q}, t)$: it is no longer defined at equal time. This means that our measure μ needs to be modified to define integrals over the time variable.[23]

As we can see, the framework we have is ill suited to handle these cases. Therefore we cannot simply stick time in the Hamiltonian and expect everything to work out. But we can't ignore the problem either as the three

[23] Looking ahead, another way to frame the problem is that Galilean transformations do not preserve phase space volume. Lorentz transformations, on the other hand, do[16] and it is this requirement plus the kinematic assumption that will recover them later.

cases outlined are common situations to study. So we need to go back to our definitions and amend them properly.

VI.B. Complete state space

The first thing to do is to amend our definition of state space to include all states at all times. It may seem like we are extending our definition of state and state space but, at a closer look, we are not. Since we defined state as a physical configuration at a particular time, the set of all states (i.e. the state space) is more accurately defined as the set of all configurations at all times. In the previous sections we tacitly simplified the problem by ignoring time, which made it easier to study the time independent case. This approach is common in physics and engineering, which is why it fits the standard names, but this is also the source of the above problems.

We'll therefore call *complete state space* of particle states the set of all configurations at all times and we'll indicate it by $\bar{\mathcal{S}}$. In that space, a state $\mathfrak{s} \in \bar{\mathcal{S}}$ will be given by (q^i, p_i, t), making $\bar{\mathcal{S}}$ a $2n+1$ dimensional manifold. A region at equal time will be a hypersurface $\mathcal{S}_{t=t_0} \subset \bar{\mathcal{S}}$ that includes all possible configurations at a particular time. Such hypersurface will need to cut across all evolutions once and only once, therefore not all functions $t : \bar{\mathcal{S}} \to \mathbb{R}$ are suitable time variables. A time evolution is a line $\bar{\mathfrak{s}} \subset \bar{\mathcal{S}}$ that passes once through each equal time surfaces. Specifying an evolution map means giving a set of lines such that each state is traversed once and only once.

Proposition VI.1. The complete state space $\bar{\mathcal{S}}$ for the particles of a classical material is a $2n+1$ dimensional differentiable manifold. The state space $\mathcal{S}_{t=t_0}$ at a particular time is a hypersurface of $\bar{\mathcal{S}}$.

Justification. We claim $\bar{\mathcal{S}}$ is a $2n+1$ dimensional differentiable manifold. Locally $\bar{\mathcal{S}} \supseteq U \cong T^*\mathcal{Q} \times \mathbb{R}$ as a state $\mathfrak{s} \in U$ exists for each configuration and each time instant. Locally $\bar{\mathcal{S}} \supseteq U \cong \mathbb{R}^{2n+1}$. $\bar{\mathcal{S}}$ is a $2n+1$ dimensional manifold. $T^*\mathcal{Q}$ is a differentiable manifold. Time evolution is a diffeomorphism by V.6 therefore derivatives between state variables and time variables always exist. $\bar{\mathcal{S}}$ is a differentiable manifold.

We claim $\mathcal{S}_{t=t_0}$ is a hypersurface of $\bar{\mathcal{S}}$. Let $t : \bar{\mathcal{S}} \to \mathbb{R}$ be a time variable. $\mathcal{S}_{t=t_0} = \{\mathfrak{s} \in \bar{\mathcal{S}} \mid t(\mathfrak{s}) = t_0\}$ is a level set. t is monotonic as it is a time variable. t has no critical points. t_0 is a regular value. $\mathcal{S}_{t=t_0}$ is an embedded submanifold of co-dimension 1. $\mathcal{S}_{t=t_0}$ is a hypersurface. □

For the composite state we need to take a different approach. A distribution of material over states at equal time $\rho_c : \mathcal{S}_{t=t_0} \to \mathbb{R}$ will depend on the choice of time variable, therefore the complete state space will not be useful as it will not be time invariant. What we are really interested in is the *state evolution space* of a classical material, which we'll indicate by $\bar{\mathcal{C}}$. To each state evolution $\bar{c} \in \bar{\mathcal{C}}$ will correspond a distribution $\rho_{\bar{c}} : \bar{\mathcal{S}} \to \mathbb{R}$ over all particle states at all times. This distribution will be differentiable: it's differentiable over all equal time hypersurfaces and time evolution is differentiable. The distribution, though, is not integrable over the complete state space $\bar{\mathcal{S}}$. Such integration will operate over time as well, giving the amount of material multiplied by time. If time is taken to be infinite the value is not finite. The distribution, instead, must be integrable over all equal time hypersurfaces: no matter what time variable we choose, the amount of material found at a particular time has to be finite.

Proposition VI.2. The state evolution space $\bar{\mathcal{C}}$ for a classical material is isomorphic to the space of differentiable functions that are Lebesgue integrable on any equal time hypersurface. That is $\bar{\mathcal{C}} \cong C^1(\bar{\mathcal{S}}) \cap L^1_t(\bar{\mathcal{S}}, \mu)$ where $L^1_t(\bar{\mathcal{S}}, \mu) = \{\rho : \bar{\mathcal{S}} \to \mathbb{R} \mid \int_{\mathcal{S}_{t=t_0}} |\rho| d\mu < \infty \; \forall t, t_0\}$.

Justification. We claim $\bar{\mathcal{C}} \cong G \subseteq C^1(\bar{\mathcal{S}}) \cap L^1_t(\bar{\mathcal{S}}, \mu)$. Let $\rho_{\bar{c}} : \bar{\mathcal{S}} \to \mathbb{R}$ be the distribution of material associated with an evolution $\bar{c} \in \bar{\mathcal{C}}$. Let $t : \bar{\mathcal{S}} \to \mathbb{R}$ be a time variable. The restriction of $\rho_{\bar{c}}$ over an arbitrary level set $\mathcal{S}_{t=t_0}$ is continuous and integrable by V.4. $\rho_{\bar{c}} \in L^1_t(\bar{\mathcal{S}}, \mu)$. Time evolution is a diffeomorphism by V.7. $\rho_{\bar{c}}$ is differentiable along the time variable as well as the state variables. $\rho_{\bar{c}} \in C^1(\bar{\mathcal{S}})$.

We claim $\bar{\mathcal{C}} \cong C^1(\bar{\mathcal{S}}) \cap L^1_t(\bar{\mathcal{S}}, \mu)$. As in V.4, if two distributions are different there exists an outcome that can tell them apart, therefore they represent two distinct physical evolutions. □

Before proceeding, it's useful to get a better intuition for the geometry of these spaces. Note that the transported state variables are perfect to label the particle state evolution $\bar{\mathfrak{s}} \subset \bar{\mathcal{S}}$: as they do not change during the evolution we can take them to label not just the values at a particular time, but the evolutions themselves. The distribution is transported deterministically and reversibly over those lines, so the value of the distribution will be constant along each particle state evolution $\bar{\mathfrak{s}}$. To identify the full evolution, then, we just need to specify the distribution only on a region that cuts across all particle evolutions, i.e. a hypersurface at constant time.

VI.C. Relativistic cotangent bundle

We now want to extend the symplectic form so we can count state and possibilities over the complete state space $\bar{\mathcal{S}}$. The first thing to do is to complete our n-dimensional $\mathcal{Q}$, the manifold that defines our units, to include the time variable t. This, again, highlights our previous oversight: we didn't include time within the quantities needed to specify a physical configuration at a particular time. We call $\mathcal{M}$ such a space which is a manifold of dimension $n+1$: state variables and the time variable are independent (i.e. any combination of state and time variables is valid).

As we said before, the time variable is not simply another state variable: it doesn't identify extra configurations and it's the variable on which we have defined deterministic and reversible evolution. When changing variables, we still need to make sure that deterministic and reversible motion can be defined on the new time variable. Transformations like $\hat{t} = q,\ \hat{q} = t$ (exchanging time with another state variable) or $\hat{t} = \sqrt{t^2 + q^2},\ \hat{q} = \tan^{-1}(t/q)$ (polar coordinates between time and state variable) clearly do not guarantee deterministic and reversible motion over the new time variable. To be meaningful, the change of time variable must preserve time ordering therefore the transformation has to be strictly monotonic between the two time variables. On the other hand, the change of state variables must preserve the ability to uniquely identify states at each instant in time. In this sense, time is not just like another state variable.[24]

As in the time independent case, we are interested in invariant densities which are defined on co-vectors. The co-vector $k_i dq^i$ is completed with a time component $\bar{\omega} dt$.[25] As each state variable q^i has a conjugate quantity $p_i \equiv \hbar k$, the time variable t will have a conjugate quantity $E \equiv \hbar\bar{\omega}$. We call the combination of (t, E) the temporal degree of freedom. As time is special, so is the temporal degree of freedom, which is treated differently when defining a distribution ρ_c and therefore by the measure μ and the symplectic form ω.

Most of all, the temporal degree of freedom is not an independent degree of freedom because E cannot be an independent variable. It cannot add any configurations, as those are defined by $T^*\mathcal{Q}$ only, and cannot add any time instants, as those are defined by t alone. Therefore there must exist a constraint such that, locally, $E = E(t, q^i, p_i)$. The complete state space $\bar{\mathcal{S}}$ is a hypersurface of $T^*\mathcal{M}$.

As the temporal degree of freedom (t, E) is not independent of the other d.o.f. (q^i, p_i); it is not orthogonal to them in $T^*\mathcal{M}$. States are defined on the plane where (q, p) (maximally) change. This is not the plane of constant (t, E) (they are not orthogonal) where the area given by $dq \wedge dp$ is defined. It is the plane perpendicular to constant (q, p). And the plane of constant (q, p) is where the area given by $dt \wedge dE$ is defined. That is: the plane where we can properly count states and define our symplectic form ω is perpendicular to $dt \wedge dE$. Intuitively, states are defined independently of time, therefore they are defined on a surface perpendicular to the temporal d.o.f.

We have a right triangle-like relationship between the plane where ω is defined and its projections on the planes

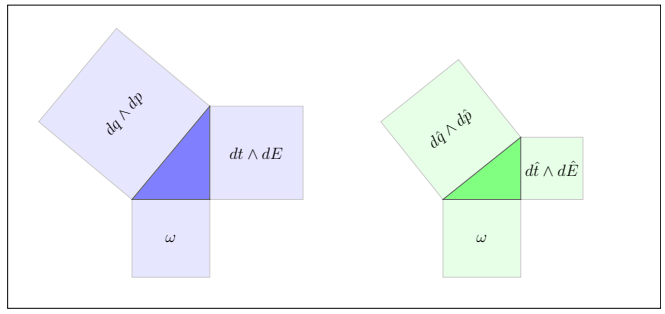

FIG. 3. The symplectic form ω and its components along temporal and standard degrees of freedom. The two d.o.f. are not independent and therefore not orthogonal. The temporal d.o.f. is orthogonal to the symplectic form as states are defined (and counted) independently of time. Compare with FIG. 2. Note: this is a 2D conceptualization of 4D space.

defined by each d.o.f., similar to the multiple d.o.f.:

$$\begin{aligned} \text{multiple d.o.f.} \quad & dq^1 \wedge dp_1 + dq^2 \wedge dp_2 = \omega \\ \text{temporal d.o.f.} \quad & dt \wedge dE + \omega = dq \wedge dp \end{aligned}$$

But in the previous case, the right angle was between the two independent d.o.f.. In this case, the right angle is between ω and the plane of constant (q, p) where $dt \wedge dE$ is defined. We rewrite it as $dq \wedge dp - dt \wedge dE = \omega$. This corresponds to the Minkowski product across d.o.f. and the vector product within. In terms of unified state variables $\xi^a \equiv \{t, q^i, E, p_i\}$ we have:

$$\omega_{ab} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \otimes \begin{bmatrix} -1 & 0 \\ 0 & I_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & I_n \\ 1 & 0 & 0 & 0 \\ 0 & -I_n & 0 & 0 \end{bmatrix}$$

The symplectic form ω is still the canonical two-form expressed with the mathematically non-canonical, but physically meaningful, variable E. This allows us to better understand its physical meaning. ω allows us to count the possibilities within a degree of freedom, adjusting it on $T^*\mathcal{M}$ to avoid the counting problems introduced by time variable changes.

The symplectic form ω also allows us to distinguish temporal from standard d.o.f. Given a two-dimensional surface $U \subset T^*\mathcal{M}$, how can we tell if it should be charted by a (t, E) or whether we can define a distribution on it? Consider $\int_U \omega$. The result will be positive if the integration is over a standard degree of freedom, and it will be negative if the integration is over the temporal degree of freedom. If the contribution in every subregion of U is positive, then the whole U is always oriented along a standard degree of freedom.

The invariance of ω also prevents mixing between standard and temporal d.o.f. The Euclidean signature across standard d.o.f. allows us to move ranges of possibilities

[24] In relativity, for example, it is common to use foliations and the $3 + 1$ formalism to recover the special character of time when performing calculations or to gain better physical insight[17].

[25] $\bar{\omega}$ is the classical analogue of the wave frequency. We use $\bar{\omega}$ to distinguish from the symplectic form ω.

from one d.o.f. to the other while conserving the total: the generators of rotation are in fact divergence free. The Minkowski signature, instead, only allows hyperbolic rotation (i.e. Lorentz boosts) across standard and temporal d.o.f., whose generators are curl free. Therefore, as required, we cannot transform a distribution over a standard d.o.f. into a distribution over a temporal one.

We can capture the above discussion by stating that the complete state space $\bar{\mathcal{S}}$ is a hypersurface of the symplectic manifold $(T^*\mathcal{M}, \omega)$, where $\omega \equiv \sum dq^i \wedge dp_i - dt \wedge dE$. And if we set $q^\alpha = (t, q^i)$ and $p_\alpha = (-E, p_i)$, we can write $\omega \equiv \sum dq^\alpha \wedge dp_\alpha$. This is the manifold that allows us to define time-and-state-variable-invariant densities. The symplectic form ω allows us to measure the count of possibilities as in the time independent case.

Proposition VI.3. The complete state space $\bar{\mathcal{S}}$ for the particles of a classical material is a hypersurface of the symplectic manifold formed by a cotangent bundle $T^*\mathcal{M}$ equipped with the canonical two-form $\omega = \sum dq^i \wedge dp_i - dt \wedge dE = \sum dq^\alpha \wedge dp_\alpha$.

Justification. We claim the space of physical objects that allows time-and-state-variable-invariant densities is a symplectic manifold $(T^*\mathcal{M}, \omega)$ where $\omega = \sum dq^i \wedge dp_i - dt \wedge dE$. Let $\mathcal{M}$ be the manifold that defines the unit system including time. Locally $\mathcal{M} \cong \mathcal{Q} \times \mathbb{R}$ as time and state variables are independent quantities. As in V.5, invariant densities are defined on the symplectic manifold $(T^*\mathcal{M}, \omega)$. Let $\bar{\omega}$ be the co-vector component associated with t. Let $E \equiv \hbar\bar{\omega}$. Let ω be the symplect form. The components of ω across different d.o.f. must be zero as they are not invariant under change of units. $\omega = \sum dq^i \wedge dp_i \pm dt \wedge dE$. The $+$ case is excluded as the temporal degree is not an independent d.o.f.: $\omega = \sum dq^i \wedge dp_i - dt \wedge dE$.

We claim $\bar{\mathcal{S}}$ is a hypersurface of $(T^*\mathcal{M}, \omega)$. $\bar{\mathcal{S}} \subseteq (T^*\mathcal{M}, \omega)$ as states are physical objects that allow time-and-state-variable-invariant densities. Consider the map $f : T^*\mathcal{M} \to \bar{\mathcal{S}} \mid (q^i, p_i, t, E) \mapsto (q^i, p_i, t)$. f is an open map. The topology of $\bar{\mathcal{S}}$ is the subspace topology. Let $\rho : T^*\mathcal{M} \to \mathbb{R}$ be an invariant density. ρ is differentiable by V.4. Its restriction on $\bar{\mathcal{S}}$ is also an invariant distribution and therefore differentiable. The inclusion map $\bar{\mathcal{S}} \hookrightarrow T^*\mathcal{M}$ is a smooth embedding. $\bar{\mathcal{S}}$ is an embedded submanifold of co-dimension 1. $\bar{\mathcal{S}}$ is a hypersurface. □

VI.D. Relativistic Hamiltonian mechanics

As we are using time as a variable to identify states, we will use a different quantity as the parameter for the evolution. The trajectory of a particle in the extended phase space will be given by the evolved variables $\xi^a(s)$ in terms of a parameter s. As before, deterministic and reversible evolution will preserve ω, as the possibility count on each independent d.o.f. will be conserved. Mathematically, deterministic and reversible evolution is a self-symplectomorphism.

Proposition VI.4. A time-dependent deterministic and reversible evolution map for the particles of a classical material is a self-symplectomorphism on $T^*\mathcal{M}$. That is: $\mathcal{T}_{\Delta s} : T^*\mathcal{M} \to T^*\mathcal{M}$ such that $\mathcal{T}^*_{\Delta s}\omega = \omega$ where $\mathcal{T}^*_{\Delta s}$ is the pullback of $\mathcal{T}_{\Delta s}$.

Justification. We claim $\mathcal{T}_{\Delta s}$ is a self-homeomorphism on $T^*\mathcal{M}$. The complete state space for particles of a classical material is $\bar{\mathcal{S}} \subset T^*\mathcal{M}$ by VI.3. $\mathcal{T}_{\Delta s}$ is a deterministic and reversible evolution map and by IV.3 is a self-homeomorphism. We can extend the map over $T^*\mathcal{M}$.

We claim $\mathcal{T}_{\Delta s}$ is a symplectomorphism on $(T^*\mathcal{M}, \omega)$. The distribution on final states must be defined. The Jacobian for $\mathcal{T}_{\Delta s}$ exists and is non-zero. $\mathcal{T}_{\Delta s}$ is a diffeomorphism. A deterministic and reversible process conserves the number of states and possibilities. ω is the two-form that returns the count of possibilities. ω is invariant under deterministic and reversible evolution. $\mathcal{T}_{\Delta s}$ is a symplectomorphism by definition. □

If we assume continuous time evolution we have the following:

Proposition VI.5. A continuous time-dependent deterministic and reversible process for the particles of a classical material admits an invariant Hamiltonian $\mathcal{H} : T^*\mathcal{M} \to \mathbb{R}$ that allows us to write the laws of evolution as

$$\begin{aligned} d_s t &= -\partial_E \mathcal{H} \\ d_s E &= \partial_t \mathcal{H} \\ d_s q^i &= \partial_{p_i} \mathcal{H} \\ d_s p_i &= -\partial_{q^i} \mathcal{H} \end{aligned}$$

Justification. We claim the vector field $S \in T(T^*\mathcal{M})$ for the infinitesimal displacement $S^a = d_s\xi^a$ admits a potential $\mathcal{H}$ such that $S^a\omega_{ab} = \partial_b\mathcal{H}$. The map for infinitesimal evolution $\mathcal{T}_{ds}$ is an infinitesimal self-symplectomorphism by VI.4. By A.3 the infinitesimal displacement S admits a potential $\mathcal{H}$ such that $S^a\omega_{ab} = \partial_b\mathcal{H}$

We claim the state variables evolve according to the extended Hamilton equations. $S^a\omega_{ab} = d_s\xi^a\omega_{ab} = \partial_b\mathcal{H}$. For $a = 0$ we have $d_s t\,(-1) = \partial_E\mathcal{H}$. For $a = \{1, ..., n\}$ we have $d_s q^i\,(+1) = \partial_{p^i}\mathcal{H}$. For $a = n+1$ we have $d_s E\,(+1) = \partial_t\mathcal{H}$. For $a = \{n+2, ..., 2n+1\}$ we have $d_s p_i\,(-1) = \partial_{q^i}\mathcal{H}$. □

We recognize the equations as those for Hamiltonian mechanics on the extended phase space[18–20].

The trajectory in time $t(s)$ is of particular importance. Since the evolution is deterministic and reversible, for each value of s we need to have one and only one value of t. Therefore $t(s)$ is invertible, strictly monotonic and $d_s t$ along a trajectory cannot change sign. This means there are two classes of states: the ones where the parametrization s is aligned with time t, which we call standard states, and the ones where the parametrization s is anti-aligned with time t, which we call anti-states. Let us call $\lambda : T^*\mathcal{M} \to \mathbb{R}$ the function $\lambda(\xi^a) \mapsto d_s t|_{\xi^a}$ that returns the change of t along s. $\lambda > 0$ for all standard

states while $\lambda < 0$ for all anti-states. Note that since the parametrization is conventional and can be changed to $s' = -s$, what we call standard and anti-states is also conventional. What is physical and not conventional, though, is that standard and anti-states cannot be connected by deterministic and reversible evolution.[26]

Proposition VI.6. Let $\mathcal{T}_{\Delta s} : T^*\mathcal{M} \to T^*\mathcal{M}$ be the time dependent deterministic and reversible evolution map for the particles of a classical material. The map partitions the extended state space into standard states, those connected by a trajectory where $d_s t > 0$, and anti-states, those connected by a trajectory where $d_s t < 0$.

Justification. We claim $t(s)$ is strictly monotonic. The motion is deterministic and reversible. At each value of t there must be only one possible state: if there is more than one state the motion is non-deterministic (one state is not enough to identify more than one state), if there is no state the motion is non-reversible (no state cannot identify a state). $t(s)$ is invertible. $t(s)$ is strictly monotonic.

We claim $d_s t$ partitions the space. Let $\mathscr{s}_1, \mathscr{s}_2 \in T^*\mathcal{M}$ be two physical states connected by deterministic and reversible evolution. $\text{sign}(d_s t(\mathscr{s}_1)) = \text{sign}(d_s t(\mathscr{s}_2))$. Let $U \equiv \{\mathscr{s} \in T^*\mathcal{M} \mid d_s t(\mathscr{s}) > 0\}$ and $V \equiv \{\mathscr{s} \in T^*\mathcal{M} \mid d_s t(\mathscr{s}) < 0\}$. $U \cap V = \varnothing$. Let $\gamma : [s_0, s_1] \to T^*\mathcal{M}$ the trajectory given by deterministic and reversible evolution, where $[s_0, s_1]$ is the range of the parametrization. Either $\gamma([s_0, s_1]) \subseteq U$ or $\gamma([s_0, s_1]) \subseteq V$. □

With the invariant Hamiltonian we are able to write the equations of motion for any choice of time and state variables. For example: $d_t q^i = d_t s\, d_s q^i = d_s q^i / d_s t = -\partial_{p_i}\mathcal{H}/\partial_E\mathcal{H}$. These equations must be the same as the ones given by a standard Hamiltonian in those coordinates. That is: $d_t q^i = \partial_{p_i} H = -\partial_{p_i}\mathcal{H}/\partial_E\mathcal{H}$. The partial derivatives of H and $\mathcal{H}$ are related, and therefore the functions themselves are related. Working through the math, we find that the most general relationship is of the form: $\mathcal{H} = \lambda(t, q^i, E, p_i)\,(H(t, q^i, p_i) - E - \hbar(t))$ where $\lambda = d_s t$, $\hbar(t)$ is an arbitrary function and, most importantly, $E = H(t, q^i, p_i) + \hbar(t)$.

The arbitrary function $\hbar(t)$ has no physical consequence: it is just a constant added to the Hamiltonian. We can either set it to zero or simply redefine the Hamiltonian to include it. Therefore we are left with $E = H(t, q^i, p_i)$. In other words: E is the value of the Hamiltonian, the energy of the system. We have found the constraint we discussed when introducing VI.3.

Note that $\mathcal{H} = 0$ for all states $\mathscr{s} \in \bar{\mathcal{S}} \subset T^*\mathcal{M}$, since $\mathcal{H} = \lambda(H - E)$ and $E = H$. For all states we also have $\lambda \neq 0$. Therefore $\mathcal{H} = 0$ only because $E = H$. We can extend $\mathcal{H}$ on all of $T^*\mathcal{M}$ such that $\mathcal{H} \neq 0$ outside of the complete state space $\bar{\mathcal{S}}$. This way we can directly use the constraint $\mathcal{H} = 0$ to identify $\bar{\mathcal{S}}$ as a subspace of $T^*\mathcal{M}$.

[26] This represents the classical analogue for quantum anti-particle states.

This way $\mathcal{H}$ gives both the constraint $\mathcal{H} = 0$ to identify states and the map $d_s\xi^a = \omega^{ab}\partial_b\mathcal{H}$ to identify their evolution.

Proposition VI.7. The form of the invariant Hamiltonian is given by $\mathcal{H} = \lambda(t, q^i, E, p_i)\,(H(t, q^i, p_i) - E)$ where $H(t, q^i, p_i)$ is the standard Hamiltonian at each instant t, the conjugate time variable $E = H(t, q^i, p_i)$ represents the value of the Hamiltonian, and $\lambda : \bar{\mathcal{S}} \to \mathbb{R}$ where $\lambda(\xi^a) = d_s t$. $\bar{\mathcal{S}}$ is the level set for $\mathcal{H} = 0$.

Justification. We claim the form of the extended Hamiltonian is $\mathcal{H} = \lambda(t, E, q^i, p_i)\,(H(t, q^i, p_i) - E)$. Let $\mathcal{H}$ be the invariant Hamiltonian and H be the Hamiltonian for a particular choice of time and state variables. We have $d_s q^i = \partial_{p_i}\mathcal{H} = d_s t\, d_t q^i = -\lambda\,\partial_{p_i} H$ where $\lambda : T^*\mathcal{M} \to \mathbb{R}$ such that $\lambda \equiv d_s t$. $d_s p_i = -\partial_{q^i}\mathcal{H} = d_s t\, d_t p_i = \lambda\,\partial_{q^i} H$. As $\lambda \neq 0$ for all physical states, we can set $\mathcal{H} = \lambda(t, q^i, E, p_i)\,(H(t, q^i, p_i) - E + \hbar(t, q^i, E, p_i))$ without loss of generality. We have $\partial_{q^i}\lambda\,(H - E + \hbar) + \lambda\,(\partial_{q^i} H + \partial_{q^i}\hbar) = \lambda\,\partial_{q^i} H$. This holds for all choices of t. In particular for $t = s$, $\lambda\,(\partial_{q^i} H + \partial_{q^i}\hbar) = \lambda\,\partial_{q^i} H$. $\partial_{q^i}\hbar = 0$. $\hbar$ is not a function of q^i. Repeat the same procedure for p_i. $\hbar$ is not a function of p_i. $\lambda = d_s t = -\partial_E\mathcal{H} = -\partial_E\lambda(H - E + \hbar) - \lambda\,(-1 + \partial_E\hbar)$. In particular for $t = s$, $\lambda = -\lambda(-1 + \partial_E\hbar)$. $\partial_E\hbar = 0$. $\hbar$ is not a function of E. $\hbar$ is only a function of t. We can redefine $H = H + \hbar$ as the result is still a function of q^i, p_i, t with the same equations of motion. $\mathcal{H} = \lambda(H - E)$.

We claim $E = H(t, q^i, p_i)$ for physical states. $\lambda = d_s t = -\partial_E\mathcal{H} = -\partial_E\lambda(H - E) - \lambda(-1)$ regardless of the choice of t and therefore for all λ. $H - E = 0$.

We claim $\bar{\mathcal{S}}$ is the level set for $\mathcal{H} = 0$. $\mathcal{H} = \lambda(H - E) = 0$ over $\bar{\mathcal{S}}$. We can extend $\mathcal{H}$ over $T^*\mathcal{M}$ such that $\mathcal{H} \neq 0$ over $T^*\mathcal{M} \setminus \bar{\mathcal{S}}$. $\bar{\mathcal{S}}$ is the level set for $\mathcal{H} = 0$. □

As an example, the Hamiltonian and invariant Hamiltonian for a relativistic free particle are:

$$\begin{aligned} H &= c\sqrt{m^2c^2 + p^i p_i} \\ \mathcal{H} &= \frac{1}{2m}(p^i p_i - (E/c)^2 + m^2c^2) \\ &= \frac{1}{2mc^2}(c\sqrt{m^2c^2 + p^i p_i} + E)(c\sqrt{m^2c^2 + p^i p_i} - E) \\ \lambda &= \frac{1}{2mc^2}(c\sqrt{m^2c^2 + p^i p_i} + E) \end{aligned} \tag{VI.8}$$

Therefore we can see that the invariant Hamiltonian has the form $\mathcal{H} = \lambda(H - E)$. In the next section we'll derive the above expression and look more closely at the case of particles under potential forces.

Now that we have characterized the motion of the particles, we should also characterize the evolution of composite states.

Suppose we have a distribution of material $\rho_{\bar{c}}(s)$ associated with an evolution $\bar{c}$. Its support, the region where the density is non-zero, is within $\bar{\mathcal{S}}$. Therefore $\mathcal{H}\rho_{\bar{c}}(s) = 0$ since $\mathcal{H} = 0$ in $\bar{\mathcal{S}}$ where $\rho_{\bar{c}}(s) \neq 0$. As the density is transported over each trajectory, its value does not change over s. Therefore we have $d_s\rho_{\bar{c}} = d_s\xi^a\partial_a\rho_{\bar{c}} =$

$\omega^{ab}\partial_b\mathcal{H}\partial_a\rho_{\bar{c}} = \{\rho_{\bar{c}}, \mathcal{H}\} = 0$ where $\{f, g\} = \partial_a f\omega^{ab}\partial_b g$ is the Poisson bracket.

Proposition VI.9. Let $\rho_{\bar{c}}$ be the distribution of material associated with the evolution $\bar{c}$. $\mathcal{H}\rho_{\bar{c}} = 0$ and $d_s\rho_{\bar{c}} = \{\rho_{\bar{c}}, \mathcal{H}\} = 0$.

Justification. We claim $\mathcal{H}\rho_{\bar{c}} = 0\ \forall\bar{c} \in \bar{\mathcal{C}}$. Let $\rho_{\bar{c}} : T^*\mathcal{M} \to \mathbb{R}$ be the distribution associated with an evolution $\bar{c}$. $supp(\rho_{\bar{c}}) \subseteq \bar{\mathcal{S}}$ as the distribution is defined on the complete state space of the particles. $supp(\mathcal{H}) = T^*\mathcal{M} \setminus \bar{\mathcal{S}}$ by VI.7. $supp(\mathcal{H}\rho_c) = supp(\mathcal{H}) \cap supp(\rho_c) = \varnothing$.

We claim $d_s\rho_{\bar{c}} = \{\rho_{\bar{c}}, \mathcal{H}\} = 0$. $d_s\rho_{\bar{c}} = \partial_a\rho_{\bar{c}}d_s\xi^a = \partial_a\rho_{\bar{c}}\omega^{ab}\partial_b\mathcal{H} = \{\rho_{\bar{c}}, \mathcal{H}\}$. Let $\gamma : \mathbb{R} \to T^*\mathcal{M}$ be a parametrization for a particle evolution $\bar{\jmath}$. Then $\rho_{\bar{c}}(\lambda(s)) = \rho_{\bar{c}}(\lambda(s+\Delta s))$ as the material does not change over the particle evolution. $d_s\rho_{\bar{c}} = 0$.

□

As we have derived relativistic motion in an unusual way, a few comments are in order.

As mentioned at the beginning of the section, no new assumption was required. Relativistic Hamiltonian mechanics is simply the correct way of handling time dependent motion. As such, there was no mention of the speed of light c, $\mathcal{M}$ being a pseudo-Riemannian manifold or its metric g. These will be introduced in the next section with the kinematic equivalence assumption. In light of this, it is probably better to consider relativistic mechanics as a feature of the geometry of the complete state space more than the geometry of space-time. Also note that this feature arises from the use of the standard topology on $\mathbb{R}^n$ and the need for invariant densities. It would not have arisen with the use of a discrete topology.

We have also seen that anti-particle states, generally associated with quantum field theory, are actually a feature of relativistic Hamiltonian mechanics. This is usually missed because in classical mechanics only the standard Hamiltonian is typically used. Note how the invariant Hamiltonian in VI.8 is quadratic in E, which can have both positive and negative solutions given the constraint $\mathcal{H} = 0$.

Another important feature is that this formulation of relativistic Hamiltonian mechanics is formally very close to the quantum case. Note how $\mathcal{H}\rho = 0$ and $\mathcal{H}\,|\psi\rangle = 0$ are formally equivalent. If we apply to the invariant Hamiltonian in VI.8 the usual formal classical/quantum substitution, we have the Klein-Gordon equation using space-like convention rescaled by a factor of $2m$. As we'll see later, to this parallel in the mathematical formalism corresponds a parallel in the physical description.

As a final thought, we note how the physics maps less elegantly to the math in the time dependent case. The function space for $\bar{\mathcal{C}}$ is not simply the Lebesgue integrable differentiable functions, but those that are integrable on particular hypersurfaces. The complete state space $\bar{\mathcal{S}}$ is a hypersurface of $T^*\mathcal{M}$. The invariant Hamiltonian is needed to identify that subspace. This may be a hint that a better formulation is lurking, one where we consider the particle evolutions $\bar{\jmath}$ as the primary objects and label those (instead of the states). The distribution of material would simply be a distribution over the particle evolutions, a Lebesgue integrable differentiable function. Yet, the need to be able to describe the evolution in terms of physically meaningful state variables will most likely reimpose the extra structure.

VII. KINEMATICS AND MINIMAL MATERIALS

We will now focus our attention on the simplest possible classical material: one that is fully described by its motion in space, i.e. the internal dynamics can be ignored. What we'll find is that the dynamics of its infinitesimal parts follows classical Lagrangian mechanics. Furthermore, the Hamiltonian of the system is constrained to the one of a free particle under scalar and vector potential forces.

VII.A. Kinematic equivalence

The link between kinematics, the study of motion through quantities like velocity and acceleration, and dynamics, the study of state evolution and its causes, is crucial in physics. In Newtonian physics, $\vec{F} = m\vec{a}$ provides such a link. In Hamiltonian (and Lagrangian) mechanics, the link is implicitly given by the choice of the Hamiltonian (or Lagrangian).

Our aim is to make that link more conceptually crisp and explicit with the following assumption:

Assumption III (Kinematic equivalence)**.** *For the system under investigation, studying its kinematics (i.e. trajectories in space) is equivalent to studying its dynamics (i.e. states and their evolution).*

Rationale. The idea is that the system internal dynamics has no effect on the overall motion and therefore can be disregarded. Consider the motion of our cannonball under gravitational and inertial forces: its temperature or precise chemical composition does not affect the overall trajectory. We can assume kinematic equivalence. Suppose we now add magnetic forces to our consideration. The magnetization of the cannonball may now affect the motion and therefore we cannot always reconstruct the full state by simply looking at the trajectory. We cannot assume kinematic equivalence.

Kinematic equivalence can't therefore be assumed for all systems. Yet, the assumption is fundamental because the particles of any material, during their evolution, will *at least* describe a trajectory in space and what we learn about the link between kinematic and dynamical quantities applies to all materials, even the ones where the kinematic assumption does not hold.

The first natural question to ask is: why does state evolution of a particle of a classical material always describe a trajectory in space? It's because spatial extent

is the one way in which all materials can be decomposed. As we divide materials into smaller pieces, physical location is one state variable we can assign to the particles of all materials. If we assume kinematic equivalence, the kinematic variables will be enough to identify a particle state. In such a case, we will call the material *minimal* to highlight the idea that the state is described by the smallest possible number of d.o.f..

When the system is not minimal, the state of the particles will be described by some additional independent state variables. The overall dynamics will be the spatial d.o.f. coupled with the internal d.o.f.. Therefore by studying a minimal material, we are also studying a component of the dynamics that is present in all materials. In other words: any material is an extension to a minimal material.

It should also be noted that, once we reduced particles to be infinitesimal at a point, the rest of the description must be a local object invariant under coordinate transformations. This means that the remaining internal state must be identified by scalar, vector or tensor quantities. This may be used in the future to justify the extension of this work to field theories. □

VII.B. Space-time manifold and Lagrangian mechanics

As we need to describe trajectories, we introduce $\mathcal{M}$ as the $n+1$ dimensional manifold that defines physical space and time. A point in $\mathcal{M}$ will be identified by $x^\alpha = (t, x^i)$, its time and space coordinates. A trajectory will be given by a set of functions $x^\alpha(s)$ where s is an arbitrary parametrization of the curve. What we need to understand is the link between the complete state space of the particles of a classical material and the space of all the possible trajectories under the kinematic assumption.

We first note that the system of units for the trajectories is fully defined by $\mathcal{M}$. Any trajectory can be specified once a set of coordinates x^α is chosen. Under the kinematic assumption, these units must also be sufficient to describe the dynamics. Therefore $\mathcal{M}$ has conceptually two roles: it is the space where the motion unfolds and it is the space that defines the system of units for our states. This already tells us that the state space for the particles of the material is a hypersurface of $T^*\mathcal{M}$.

Under kinematic equivalence, given a trajectory we must identify a state and vice-versa. There must be a homeomorphism between the complete state space and the space of possible trajectories. We already know what the complete state space is. We just need to find a way to label the trajectories so that we can express the correspondence.

Because of the double role of $\mathcal{M}$, we can set half of our state variables to be equal to the space-time variables: $q^\alpha = (t, q^i) = x^\alpha$. This already tells us that the trajectory is differentiable since $q^\alpha(s)$ is differentiable. The four-velocity $u^\alpha = d_s x^\alpha$ is therefore well defined and it can be used to identify trajectories. The magnitude of the four-velocity can be rescaled by changing the evolution parameter, $u^\alpha = d_s x^\alpha = d_s \hat{s}\, d_{\hat{s}} x^\alpha = d_s \hat{s}\, \hat{u}^\alpha$, and is therefore not physical. Only the direction is physical so the four-velocity gives us another n independent variables to identify trajectories. The combined position and velocity, the kinematic variables, identify $2n+1$ trajectories using a hypersurface of the tangent bundle $T\mathcal{M}$. This is exactly the dimensionality of the complete state space $\bar{\mathcal{S}} \subset T^*\mathcal{M}$. We cannot add more curves, as a homeomorphism preserves dimensionality, therefore we identified the suitable space for the trajectories of particles of a classical material under kinematic equivalence.

Being able to identify particles through either state or kinematic variables is not enough: we also need to be able to express a composite state as the distribution of particle states over position and velocity. As such, the differentiable distribution over state variables must become a differentiable distribution over kinematic variables. This tells us the transformation is differentiable.

Mathematically, the state space $\bar{\mathcal{S}} \subset T^*\mathcal{M}$ is diffeomorphic to a hypersurface of the tangent bundle $T\mathcal{M}$ of the space-time manifold.

Proposition VII.1. Let $\mathcal{M}$ be the $n+1$ dimensional space-time manifold over which trajectories are defined. The complete state space $\bar{\mathcal{S}} \subset T^*\mathcal{M}$ for the particles of a minimal classical material is diffeomorphic to a hypersurface of the tangent bundle $T\mathcal{M}$.

Justification. We claim the complete state space is a hypersurface of the cotangent bundle of the space-time manifold, $\bar{\mathcal{S}} \subset T^*\mathcal{M}$. Let $\bar{\mathfrak{s}} \in \bar{\mathcal{S}}$ be the evolution of a particle for a minimal classical material. Let $f(\bar{\mathfrak{s}}) \mapsto x^\alpha(s)$ be the function that returns the space-time trajectory associated with the evolution. f is injective as the material is minimal. $\bar{\mathfrak{s}}$ can be identified by a trajectory $x^\alpha(s)$ which can be fully described by coordinates defined on $\mathcal{M}$. Let $\mathfrak{s} \in \bar{\mathcal{S}}$ be the particle state for a minimal classical material. Let $f(\bar{\mathfrak{s}}, t) \mapsto \mathfrak{s}$ the function that returns the state at a particular time t along an evolution $\bar{\mathfrak{s}}$. Such function is a bijection as the evolution is deterministic and reversible. $\mathfrak{s}$ can be identified by an evolution $\bar{\mathfrak{s}}$ and an instant t, both fully described by coordinates defined on $\mathcal{M}$. $\mathcal{M}$ defines the unit system. $\bar{\mathcal{S}} \subset T^*\mathcal{M}$ by VI.3.

We claim there exists a homeomorphism between $\bar{\mathcal{S}} \subset T^*\mathcal{M}$ and a hypersurface of $T\mathcal{M}$. Let $\mathfrak{s} \in \bar{\mathcal{S}}$ be the particle state for a minimal classical material. Let $x^\alpha(s)$ be the evolution associated with that state. Let $t(\mathfrak{s})$ be the instant of time at which $\mathfrak{s}$ is defined. $q^\alpha(\mathfrak{s}) \equiv x^\alpha(s(t(\mathfrak{s})))$ are state variables as they are real function of the state. q^α is differentiable therefore $d_s x^\alpha$ exists. Suppose $d_s x^\alpha$ is not part of the state. Then $d_s x^\alpha = d_s q^\alpha$ must be determined by the law of evolution. $x^\alpha(s)$ is determined by the law of evolution therefore q^α are the only state variables needed from the trajectory. As the material is minimal, q^α determines the full state $\mathfrak{s}$. This is a contradiction: $\mathfrak{s}$ is not identified by just q^α. Therefore $u^\alpha = d_s x^\alpha$ is part of the state. $p_\alpha = p_\alpha(x^\alpha, u^\alpha)$. The map $(x^\alpha, u^\alpha) \mapsto (q^\alpha, p_\alpha)$

is continuous by IV.4 and is bijective as the material is minimal. It is a homemorphism between $T^*\mathcal{M}$ and $T\mathcal{M}$. $\bar{\mathcal{S}}$ is a hypersurface of $T^*\mathcal{M}$ therefore its image is a hypersurface of $T\mathcal{M}$.

We claim $(x^\alpha, u^\alpha) \mapsto (q^\alpha, p_\alpha)$ is a diffeomorphism. Let $\bar{c}$ be the evolution of a minimal classical material. Let $\rho_{\bar{c}} : \bar{\mathcal{S}} \to \mathbb{R}$ be the distribution associated with $\bar{c}$. $\rho_{\bar{c}}(\imath(x^\alpha, u^\alpha))$ as the material is minimal. $\rho_{\bar{c}}(x^\alpha, u^\alpha)$ is differentiable by the same logic expressed in V.4. $(x^\alpha, u^\alpha) \mapsto (q^\alpha, p_\alpha)$ is a diffeomorphism as it is a homeomorphism that maps a differentiable function to a differentiable function. □

This relationship allows us to write the kinematic variables in terms of the state variables. In particular we have $u^\alpha = u^\alpha(q^\alpha, p_\alpha)$. As the relationship is invertible, these functions must be monotonic in p_α. Since $\partial_{p_\alpha}\mathcal{H} = d_s q^\alpha = u^\alpha$, $\mathcal{H}$ is convex in p_α. Under such conditions we can define a Legendre transformation $\mathcal{L} = u^\alpha p_\alpha - \mathcal{H}$. Such transformation is the Lagrangian of the system and the motion is given by the Euler-Lagrange equation. For the time independent case, we define $L = v^i p_i - H$ where $v^i = d_t x^i$.

Proposition VII.2. Let $\mathcal{H}$ be an invariant Hamiltonian defined on $T^*\mathcal{M}$. Under the kinematic assumption, we can define the Legendre transform $\mathcal{L} = u^\alpha p_\alpha - \mathcal{H}$ which we call the *extended Lagrangian*.

Justification. We claim $\mathcal{H}$ is convex in p_α. By VII.1 we can write the velocity $u^\alpha = u^\alpha(q^\alpha, p_\alpha)$ as a differentiable function of the state variables. At fixed q^α, u^α must be invertible. The components u^α are monotonic in p_α. $\partial_{p_\alpha}\mathcal{H} = d_s q^\alpha = u^\alpha$. The derivatives of $\mathcal{H}$ in p_α are monotonic in p_α. $\mathcal{H}$ is convex in p_α.

We claim there exists an extended Lagrangian $\mathcal{L} = u^\alpha p_\alpha - \mathcal{H}$. $\mathcal{H}$ is convex in p_α. It admits a Legendre transform. The extended Lagrangian $\mathcal{L}$ is such transform. □

Not only must we be able to express distributions in terms of both state and kinematic variables, but integration has to carry over as well. To do that, we need to be able to count trajectories in the same way we are able to count states. Just as the symplectic form ω gives us the count of possibilities for a d.o.f. in terms of dq^α and dp_α, the ranges of the state variables, there should be a bilinear form g that gives the the count of possibilities in terms of dx^α and du^α, the ranges of the kinematic variables. Under the kinematic assumption, these counts must be consistent with each other. That is: $\omega(dq^\alpha, dp_\alpha) \propto g(dx^\alpha, du^\alpha)$.

Since both dx^α and du^α are four vector components, g is a rank two tensor on $\mathcal{M}$. To each direction in space-time dx^α is associated a direction in the tangent space of velocities du^α with which it forms a d.o.f.. Across d.o.f., that is $\alpha \neq \beta$, we have $\omega(dq^\alpha, dp_\beta) = 0 = \omega(dq^\beta, dp_\alpha)$. Therefore if dx^α and du^β do not form a d.o.f., $g(dx^\alpha, du^\beta) = 0 = g(dx^\beta, du^\alpha)$: g is reflexive. We also know that g cannot be alternating, that is $g(dx^\alpha, du^\alpha) \neq 0$: a choice of units for x^α is a choice of units for u^α and therefore they cannot belong to different d.o.f. This means g is a symmetric non-degenerate tensor. We recognize it as the metric tensor and therefore $\mathcal{M}$ is a pseudo-Riemannian manifold.

Since g is symmetric, it can be diagonalized: $\omega(dq^\alpha, dp_\beta) = dq^i \omega_{q^i p_i} dp_i + dt\omega_{tE} dE \propto g(dx^\alpha, du^\beta) = dx^i g_{ii} du^i + dt g_{00} du^0$. As $\omega_{q^i p_i} > 0$ we must have $g_{ii} > 0$. As $\omega_{tE} < 0$ we must have $g_{00} < 0$. We can further change the units along the spatial directions so that $g_{ii} = 1$: if we use the same units for the spatial coordinates, each direction will contribute the same number of possibilities. Since time must use a different unit, we set $g_{00} = -c^2$ where c^2 is the constant that allows one to convert the possibility count over the temporal d.o.f. into the possibility count of a spatial d.o.f.. Therefore, locally $g(dx^i, dx^i) = (dx^i)^2 - c^2 dt^2$ is Minkowskian.

Proposition VII.3. The space-time manifold $\mathcal{M}$ is a pseudo-Riemannian manifold $(\mathcal{M}, g)$. g has signature $(n, 1)$ and locally is $g = (dx^i)^2 - c^2 dt^2$.

Justification. We claim $(\mathcal{M}, g)$ is a pseudo-Riemannian manifold where g is a metric tensor induced by ω under kinematic equivalence. Let (dx^α, du^α) be an infinitesimal range of possibilities for the kinematic variables of the particles of a minimal classical material. There exists a map $g : T\mathcal{M} \times T\mathcal{M} \to \mathbb{R}$ such that $g(dx^\alpha, du^\alpha) \propto \omega(dq^\alpha, dp_\alpha)$. The map g is linear, since a linear combination of infinitesimal ranges leads to the linear combination of possibilities. g is a rank two tensor. Suppose $g(dx^\alpha, du^\beta) = 0$. $g(dx^\alpha, du^\beta) \propto \omega(dq^\alpha, dp_\beta) = 0 = \omega(dq^\beta, dp_\alpha) \propto g(dx^\beta, du^\alpha) = 0$. g is reflexive. Position and velocity along the same direction use related units. They cannot form independent d.o.f.. $g(dx^\alpha, du^\alpha) \neq 0$. g is not alternating. g is symmetric. g is a symmetric rank two tensor. $(\mathcal{M}, g)$ is a pseudo-Riemannian manifold.

We claim g is locally Minkowskian. Let x^α be a choice of coordinate that diagonalizes g. x^α exists as g is symmetric. $\omega(dq^\alpha, dp_\beta) = dq^i \omega_{q^i p_i} dp_i + dt\omega_{tE} dE \propto g(dx^\alpha, du^\beta) \propto dq^i g_{ii} du^i + dt g_{00} du^0$. $dq^\alpha = dx^\alpha$. We require $g(dx^\alpha, du^\beta) > 0$ when the arguments are positive. Therefore $g_{ii} > 0$ and $g_{00} < 0$. g has signature $(n, 1)$. Choosing the same unit for all dx^i and a temporal unit for dt we can choose coordinates such that locally $g = (dx^i)^2 - c^2 dt^2$. g is locally Minkowskian. □

The metric tensor g and the constant c found here are indeed the ones used in the context of special relativity. Yet, they have acquired a different meaning. g does not measure space-time distances: it counts possibilities of a d.o.f., like ω. Directions in space are orthogonal not because the angle between them is 90°, but because they identify independent d.o.f.. That is, the choice of units along one direction is independent from the choice of units along orthogonal directions. Space and time, instead, are not independent d.o.f. and the geometry between them is different. In other words: g inherits its signature from ω. The geometry of space-time is the geometry of the ranges for possible states.

This also allows us to give c a more insightful physical

meaning. It is not a velocity: the metric is defined at each point without motion. It is the constant of proportionality that maps a range of initial conditions from the time variable t to the space variable x^i. For example, once the SI is adopted, 1km would correspond to a thousand times more possible states than 1m, and, in the same way, 1sec would correspond to $299,792,458$ more possible states than 1m. Therefore when we write $x^\alpha = (ct, x^i)$ or $p_\alpha = (-E/c, p_i)$ we are really expressing everything in quantities that are proportional to the count of states and possibilities. These coordinates make the bookkeeping of number of states much easier which, as we saw, is the only thing that Hamiltonian and Lagrangian mechanics are doing.

This also allows us to understand heuristically why the speed of deterministic and reversible motion cannot be greater than c. Consider the trajectory of a point particle. Over an interval dt it will travel dx in space. Since the motion is deterministic and reversible, and therefore continuous, it will travel through all the states defined in the interval dx. It will also travel through every moment in time, which corresponds to cdt possibilities. As it moves, time will always change but not necessarily space. The number of states traveled in space cannot exceed the number of states traveled in time. Therefore we have:

$$\frac{|n_x(dx)|}{|n_t(dt)|} = \frac{|dx|}{|cdt|} \leq 1$$
$$\frac{|dx|}{|dt|} \leq c$$

Note, though, that this constraint applies only to deterministic and reversible motion. It was under that premise that we were able to show that the trajectories are continuous, that we could define ω and therefore g. Non-local quantum effects, such as the ones featured in EPR[21] experiments, are not examples of deterministic and reversible evolution. Therefore they do not have anything to violate: the maximum speed does not apply to their "motion".[27] On the other hand, only deterministic and reversible motion can carry information and can establish before/after relationships, therefore everything remains consistent.

VII.C. Massive particles under scalar and vector potential forces

The link between ω and g allows us to constrain the possible diffeomorphisms between state variables and kinematic variables. If we express the possibilities in terms of both state variables and kinematic variables we have: $dq^\alpha \wedge dp_\alpha \propto dx^\alpha du_\alpha$. Since $p_\alpha = p_\alpha(x, u)$, we have $dp_\alpha = m du_\alpha$, where m is the constant of proportionality. We recognize m as the inertial mass. Integrating the expression, we have $p_\alpha = m u_\alpha + A_\alpha(x)$ where $A_\alpha(x)$ is an arbitrary vector field. This is the most general link between velocity and conjugate momentum, which is the sum of kinetic momentum and a scalar potential.

This constrains the form of the extended Hamiltonian. $\frac{1}{m} g^{\alpha\beta}(p_\beta - A_\beta) = u^\alpha = d_s x^\alpha = d_s q^\alpha = \partial_{p_\alpha}\mathcal{H}$. Integrating we have $\mathcal{H} = \frac{1}{2m}(p_\alpha - A_\alpha) g^{\alpha\beta}(p_\beta - A_\beta) + V(x)$ where $V(x)$ is an arbitrary function. We can further constrain the Hamiltonian by choosing the natural parameterization for s (also called unit speed or arc length parametrization (ref?)) for which the modulus of the velocity is constant. We have $u^\alpha u_\alpha = -c^2 = \frac{1}{m^2}(p^\alpha - A^\alpha)(p_\alpha - A_\alpha)$. $\mathcal{H} = \frac{1}{2m}((p^\alpha - A^\alpha)(p_\alpha - A_\alpha) + m^2c^2)$. Applying the same procedure in the time independent case yields the Hamiltonian $H = \frac{1}{2m}(p^i - A^i)(p_i - A_i) + V$.

We recognize the Hamiltonian for massive particles under scalar and vector potential forces.

Proposition VII.4. The extended Hamiltonian for the particles of a classical material under the kinematic assumption is $\mathcal{H} = \frac{1}{2m}((p_\alpha - A_\alpha) g^{\alpha\beta}(p_\beta - A_\beta) + m^2c^2)$.

Justification. We claim the conjugate momentum expressed as a function of the kinematic variables is $p_\alpha = m u_\alpha + A_\alpha(x)$ where $A_\alpha \in \mathfrak{X}(\mathcal{M})$ is a differentiable vector field. $\omega(dq^\alpha, dp_\beta) = dq^\alpha \wedge dp_\alpha \propto g(dx^\alpha, du^\beta) = dx^\alpha du_\alpha$. Let m be the proportionality constant. $dq^\alpha \wedge dp_\alpha = dx^\alpha \wedge \partial_{u_\beta} p_\alpha du_\beta = m dx^\alpha du_\alpha$. $\partial_{u_\beta} p_\alpha = m\delta^\beta_\alpha$. After integration $p_\alpha = m u_\alpha + A_\alpha(x)$ where A_α is an arbitrary vector field.

We claim the Hamiltonian is $\mathcal{H} = \frac{1}{2m}((p_\alpha - A_\alpha) g^{\alpha\beta}(p_\beta - A_\beta) + m^2c^2)$. $\frac{1}{m} g^{\alpha\beta}(p_\beta - A_\beta) = u^\alpha = d_s x^\alpha = d_s q^\alpha = \partial_{p_\alpha}\mathcal{H}$. After integration $\mathcal{H} = \frac{1}{2m}(p_\alpha - A_\alpha) g^{\alpha\beta}(p_\beta - A_\beta) + V(x)$ where V is an arbitrary function. Let s be the natural parametrization such that $u^\alpha u_\alpha = -c^2$. $V = \frac{1}{2}mc^2$ as $\mathcal{H} = 0$ by VI.7. $\mathcal{H} = \frac{1}{2m}((p_\alpha - A_\alpha) g^{\alpha\beta}(p_\beta - A_\beta) + m^2c^2)$.

□

The above discussion gives the inertial mass m a new physical meaning: it is the conversion constant that tells us how to convert a range of velocity to a range of conjugate momentum with the same number of possibilities. In other words: the more massive the particles, the more states per unit velocity. Intuitively, a more massive body is harder to accelerate because it has to go through more states for the same change in velocity. Also note that the Hamiltonian excludes massless particles because the kinematic assumption does that implicitly. A massless particle always has the same velocity, c, therefore the trajectory is not enough to tell us the dynamics. The motion of the photon is not enough to tell us its energy or its momentum. Therefore it makes sense that, by assuming kinematic equivalence we derived an equation valid only for massive particles.

We have seen that the inertial mass m, the invariant speed c and the reduced Planck constant $\hbar$ all serve sim-

[27] The very notion of velocity becomes ill defined in such cases precisely because we do not have a unique well-defined evolution.

ilar purposes: convert one quantity to another preserving the number of possibilities. $dp = \hbar dk$ converts between ranges of wave number and conjugate momentum. $dx = cdt$ converts between ranges of time and space. $dp = mdu$ converts between ranges of velocity and conjugate momentum. They allow us to keep proper bookkeeping when mapping sets of states. Using natural units means measuring all these quantities through the possibility count. There is a difference among them, though: while c and $\hbar$ are simply determined by the unit system, m also depends on the type of material. That is, different homogeneous materials can have different inertial mass. The reason is that the choice of units for x^α already determines the others ($x^\alpha \to u^\alpha$ and $x^\alpha \to q^\alpha \to k_\alpha \to p_\alpha$) therefore the link between units of u^α and p_α is not a new unit conversion.

Another important thing to note is that the relationship between conjugate momentum and velocity, $p_\alpha = mu_\alpha + A_\alpha(x)$, includes an arbitrary function. As it is arbitrary, we cannot give p a well defined physical meaning in terms of the space-time trajectory. Suppose, in fact, we changed A and p while leaving u unchanged, the physical description in terms of trajectories remains the same.

We should also ask whether any choice of A actually represents a different case leading to different trajectories. The simplest and more insightful way to study this problem is to use the kinematic quantities x^α and mu_α as state variables. They are not going to be conjugates of each other but they do capture the entire physical description. Any change in A that does not affect the relationship between x^α and mu_α does not change the physics. Expressing the Hamiltonian and the Poisson brackets in terms of the kinematic variables we have:

$$\begin{aligned}
\mathcal{H} &= \frac{1}{2}mu_\alpha u^\alpha + \frac{1}{2}mc^2 \\
\{x^\alpha, x^\beta\} &= 0 \\
\{x^\alpha, mu_\beta\} &= \delta^\alpha_\beta \\
\{mu_\alpha, mu_\beta\} &= \{p_\alpha, -A_\beta\} + \{-A_\alpha, p_\beta\} \\
&= \partial_\alpha A_\beta - \partial_\beta A_\alpha = F_{\alpha\beta}
\end{aligned}$$

Note how the constraint $\mathcal{H} = 0$ simply sets the rest kinetic energy. Note how the Poisson bracket between kinetic momentum is the force tensor. The weaker the force, the more (x^i, mu_i) become independent d.o.f. The force tensor, then, tells how much (x^i, mu_i) fail to be independent d.o.f.. With these expressions, the Poisson bracket between any function of position and kinetic momentum can now be calculated. For example:

$$\begin{aligned}
d_s x^\alpha &= \{x^\alpha, \mathcal{H}\} = u^\alpha \\
d_s(mu^\alpha) &= \{mu^\alpha, \mathcal{H}\} = F^{\alpha\beta}u_\beta
\end{aligned}$$

where we recognize the covariant form of the Lorentz force.

Any change of A that leaves $F_{\alpha\beta}$ unchanged does not affect the evolution of any function of position and velocity. Such transformation is what we call a gauge transformation. Note the same group of transformations is found by fixing A and changing p such that the laws of motion don't change. In other words: A and p are defined relatively to each other. A defines the 0 for p and vice-versa. The gauge group expresses the arbitrariness of such a definition.

The extended Hamiltonian we found is quadratic in energy, therefore negative energy states are possible. These correspond to the particle anti-states mentioned in VI.6. As we fixed the natural parametrization, $u^\alpha u_\alpha = -c^2$, there are two cases: the parametrization is aligned with time, therefore $s = \tau$ where τ is proper time, or the parametrization is anti-aligned with time, therefore $s = -\tau$. The trajectory velocity $u^\alpha = \pm d_\tau x^\alpha$ is plus or minus the physical velocity depending on whether it's a standard particle state or an anti-state. Expanding the equation for the force, we have:

$$\pm d_\tau(\pm m d_\tau x^\alpha) = m d_\tau^2 x^\alpha = \pm F^\alpha_\beta d_\tau x^\beta$$

Therefore the force acts on anti-particle states in the opposite way: they act as if they have opposite charge. Note that nothing physical is actually going backwards in time, just the parametrization.

In this setting, charge conjugation C is simply changing the parametrization from s to $-s$. This will permute particle standard and anti-states as they follow the trajectories in the opposite direction. But if we apply parity P and time reversal T, changing x^α to $-x^\alpha$, we would have the same net effect: the parametrization stays the same, but it now moves backwards in space and time. As they have the same net effect, if we apply all three we go back where we started. This gives us a very intuitive classical particle mechanics analogue for the CPT theorem(ref?).

Again we want to stress just how the math takes a very simple and physically meaningful form when studying the motion in phase space using position and kinematic momentum. It is most expressive because, by using the kinematic variables of a deterministic and reversible motion of an infinitesimal amount of material, it makes use of all the assumptions about the system we are studying.

To sum up, the kinematic assumption has more inherent consequences than one would expect at first. It means the particles are identified by position and velocity and therefore the system of equations is second order. It links the space of initial conditions on $T\mathcal{M}$ with the space of states on $T^*\mathcal{M}$. Such a link allows phase space symplectic form ω to induce the space-time metric tensor g. It allows us to define Lagrangians. It constrains the motion to massive particles under scalar and vector potential forces. Moreover, it allowed us to reunderstand basic concepts, like the invariant speed c, the inertial mass m and the force tensor $F_{\alpha\beta}$ in terms of state counting and d.o.f. independence. Finally, it allowed us to construct classical equivalents for anti-particles and the CPT theorem.

We now have all the building block of classical mechanics, the consequences of three seemingly simple assumptions. We have more insight of why these blocks fit together the way they do. But we have also placed ourselves in a very good position to move forward to quantum systems by using the new physical insights we have gained.

VIII. PRELUDE TO QUANTUM

Before introducing assumption IV and deriving quantum mechanics as its consequence, we first give a semi-classical account of the conceptual picture that emerges. This allows us to build the intuition more gradually and to show that classical and quantum mechanics are not as different as one may first imagine.

We will see that, as a first approximation, quantum particles are more suitably described by classical composite states (i.e. distributions in phase space) like any other object made of a finite amount of material. We'll also see that classical distributions already possess qualitative properties that are typically associated with quantum systems, including a classical analogue of the uncertainty principle.

VIII.A. The case against point particles

As we defined them in previous sections, classical particles should not be understood as point-like objects of finite mass. It was their being infinitesimal and spread around a small neighborhood that gave us conjugate pairs of state variables and Hamiltonian evolution. Any finite amount of classical material, therefore, is described by a distribution. If we take a quantum particle, say an electron, it has a finite amount of charge/mass/energy. Therefore, to a first approximation, it must be described by a distribution as well.

The idea that an electron is not a point particle may, to some, seem preposterous at first. But we contend that it is no more preposterous than to think it point-like, described by a probability distribution that self-interacts. In fact, the idea of an electron being a distribution of material is already conceptually more compatible with two features of quantum mechanics.

The first one is that quantum particles are not associated with a well defined value for position and momentum. If we misidentify the electron with a classical particle (i.e. the infinitesimal part of a classical material), this would indeed be worthy of note and puzzlement. But if the electron is a distribution, there is no issue as a phase space distribution does not have a single value for position or momentum. The earth is, so to speak, in both North America and Europe at the same time. In fact, based on what we saw before, any object of finite mass is a distribution over a region of phase space, and discontinuous distributions are ruled out by IV.4. An electron is an object of finite mass: why should it be any different? In fact, if one claimed an electron to be point-like, he would be claiming an electron is fundamentally different than all other physical objects of finite mass/energy (which are the only ones we experimentally work with). Shouldn't *he* have the burden of proof?

The other feature is particle-wave duality. If the spread of the distribution is narrow enough for the problem at hand, the motion can be described by just the position and momentum of the center of mass, making it behave like a classical particle. If the spread is not negligible (e.g. it encounters obstacles of a smaller scale) then we have to use the full distribution. Therefore, even in this classical approximation, the electron would have this dual nature depending on the problem.

While this picture is quite reasonable, some people may reject it because they believe that a fundamental object must have no size, it must be a point. But that does not follow: lack of spatial extent and lack of substructure are independent concepts.

In fact, we can conceptually construct an infinitesimal object that is not elementary. Suppose we have dots moving on the surface of a sphere. Their motion can be expressed in terms of their polar and azimuthal angles. If we shrink the sphere, the dynamics and its description stays the same. We can collapse the whole sphere to a point: it is now a point-like object that is not elementary, as there is motion detectable along different angles of incidence.

Not only do point-like objects not have to be elementary, but it's unclear how they can be elementary. How can the whole universe be reduced to elements that have no idea how many spatial dimensions are there? Mathematically a set of points by itself does not define the dimension of a space: $\mathbb{R}^n$ is the same order of infinity for any n. One needs a topology to differentiate, which is an additional structure. It's the neighborhoods that define the dimensionality. If everything is just points, where does the topology come from?

If we abandon the idea of point-like objects of finite mass, we avoid these problems and get a picture more consistent across the different realms of physics. And we can also see that classical mechanics already has some features that are commonly ascribed to quantum mechanics.

VIII.B. Classical uncertainty principle

Suppose we have our classical electron described by a finite amount of classical material distributed over a single d.o.f. $\int \rho_c dq \wedge dp = 1$ where ρ_c is the normalized distribution. In this context, we will call *fragment* an infinitesimal part of our classical electron. The Shannon entropy[11] associated with the distribution is finite as well. $I(\rho_c) = -\int \rho_c \log(\rho_c) dq \wedge dp = I_0$. Intuitively, this value can be thought of as the amount of information

required to identify a fragment within the distribution.[28]

As the state evolves into its final state, the amount of material remains the same: $\int \rho_{\hat{c}} dq \wedge dp = 1$. The final entropy will also remain the same: $I(\rho_{\hat{c}}) = I_0$. Mathematically, because $-\rho_c \log(\rho_c)$, like the distribution itself, is just another integrable function and therefore its integral is conserved by Hamiltonian evolution. Physically, because the amount of information required to identify a fragment given the distribution has to remain the same under deterministic and reversible evolution. Therefore, in the same way that each fragment is constrained to move through states at equal energy, the composite state is constrained to move through distributions at equal entropy.

We now ask the following question: once we fix the entropy, what is the relationship between the spread of the marginal distributions along q and p? Using Lagrange multipliers, we fix the total amount of material to 1 and the total entropy I_0 and find the distribution that minimizes the product of the variance $\sigma_q^2\sigma_p^2 \equiv \int (q-\mu_q)^2 \rho_c \, dqdp \int (p-\mu_p)^2 \rho_c \, dqdp$.

$$\begin{aligned}
L =& \int (q-\mu_q)^2 \rho_c \, dqdp \int (p-\mu_p)^2 \rho_c \, dqdp \\
&+ \lambda_1 (\int \rho_c dqdp - 1) \\
&+ \lambda_2 (-\int \rho_c \ln \rho_c \, dqdp - I_0) \\
\delta L =& \int \delta\rho_c [(q-\mu_q)^2\sigma_p^2 + \sigma_q^2 (p-\mu_p)^2 + \\
& \lambda_1 - \lambda_2 \ln \rho_c - \lambda_2] dqdp = 0 \\
\lambda_2 \ln \rho_c =& \lambda_1 - \lambda_2 + (q-\mu_q)^2\sigma_p^2 + \sigma_q^2 (p-\mu_p)^2 \\
\rho_c =& e^{\frac{\lambda_1-\lambda_2}{\lambda_2}} e^{\frac{(q-\mu_q)^2\sigma_p^2}{\lambda_2}} e^{\frac{\sigma_q^2(p-\mu_p)^2}{\lambda_2}}
\end{aligned}$$

We solve the multipliers and have:

$$\begin{aligned}
\rho_c =& \frac{1}{2\pi\sigma_q\sigma_p} e^{-\frac{(q-\mu_q)^2}{2\sigma_q^2}} e^{-\frac{(p-\mu_p)^2}{2\sigma_p^2}} \\
I_0 =& \ln(2\pi\sigma_q\sigma_p) + 1
\end{aligned}$$

We find that the distribution that minimizes the spread is the product of two independent Gaussians. Recall that in quantum mechanics, the Gaussian wave packet is the state that minimizes uncertainty: this is the classical analogue.

As the distribution evolves, the entropy is conserved and therefore the product $\sigma_q^2\sigma_p^2$ can never be less than the one given by the Gaussian distribution of the same entropy. We have:

$$\sigma_q\sigma_p \geq \exp(I_0 - 1)/2\pi$$

This is strikingly similar to the Heisenberg uncertainty principle, except that the value depends on the initial entropy I_0.

Suppose, though, that while the evolution of the distribution of our electron as a whole is deterministc and reversible, each fragment is undergoing non-deterministic motion. Suppose ρ is a distribution at equilibrium of such motion. While each fragment moves randomly, the overall distribution ρ stays the same: the non-deterministic motion is effectively just reshuffling the fragments within the same distribution. The value of the entropy of the distribution, and therefore its spread, is set by the non-deterministic motion of the particles: the stronger it is, the greater the entropy.

Suppose now our classical electron undergoes quasi-static evolution. At each instant we can assume ρ to be at equilibrium. At each instant the fragments are reshuffled. At each instant the entropy of the distribution is the same one associated with the random motion. The overall evolution is still Hamiltonian, but the trajectories predicted by the framework are now just the average motion of the fragments.

If we assumed that any classical electron were affected by the same source of randomness, all fragments would exhibit a random walk of the same magnitude as defined by the information entropy. The spread of all distributions for the material would satisfy the same inequality. This gives us the classical uncertainty principle: $\sigma_q\sigma_p \geq \exp(I_\varnothing - 1)/2\pi$ where $I_\varnothing$ is the entropy associated with the source.

Such an assumption is not far fetched. The source of entropy could be identified in the environment (i.e. a minimum interaction that cannot be avoided as perfect isolation cannot be achieved) or the internal self interaction of the fragments (i.e. the evolution of the unstated part). Whatever it may be, if it is characterized by an entropy comparable to the one of the distribution, that is we are trying to localize the electron better than the source can allow, the classical assumption breaks down as the random motion of each fragment is not negligible.

The generalization to multiple d.o.f. is outside of the scope of this work. Note, though, that this result is related to the Gromov non-squeezing theorem and the symplectic camel, which are already generalized to multiple d.o.f.. It is the link with information theory that gives us further physical insight and is of particular interest given the strong connection between thermodynamics and Shannon's entropy.

This discussion shows, once again, how we can already find in classical mechanics some of the ingredients that are typically assumed to be characteristics of quantum mechanics. It has its own flavor of anti-particles, CPT theorem and uncertainty principle. These should really

[28] To be precise, Shannon's differential entropy is actually the number of bits required to identify an element of the given distribution relative to one of a uniform distribution of unitary range on all axes.

be considered general features of deterministic and reversible evolution of distributions (i.e. Hamiltonian mechanics). Quantum mechanics just inherits these traits.

The discussion also gives us an intuitive physical picture that we will develop further in the next section: a quantum particle is a well determined distribution of not well determined fragments. The distribution as a whole can be studied and evolves deterministically while the fragments do not.

IX. IRREDUCIBLE SYSTEMS

In this section we explore a different relationship between the whole and its parts. We will consider an irreducible material, one where the state and the evolution of the whole does not tell us anything about the state and the evolution of the parts. The material is still conceptually infinitesimally decomposable, made of infinitesimal parts which we call fragments. But each fragment cannot be studied independently and therefore cannot be assigned a state: the internal dynamics is undetermined. We'll find that the evolution of an irreducible material is suitably described by the standard framework of quantum mechanics.

IX.A. Quantized material

In section V we introduced the idea of a classical material, one that is homogeneous and infinitesimally reducible, which represented a way to relate the state of the whole system to the state of its parts. In this section we explore an alternative case.

We will call a *quantized material* a homogeneous material that is finitely reducible. That is, as we decompose the system we can only assign a state to finite amounts of material, which we call *quantum particles*. As the material is homogeneous, all its particles carry the same amount of material and are described by the same state space. Each quantum particle is irreducible: its state does not tell us anything about the state of its parts. While this idea should be familiar enough, it is best to iron out some details to avoid possible misconceptions.

The first possible source of confusion is between divisibility, the idea that an object can be divided into parts, and reducibility, the idea that the state of the whole is equivalent to the state of the parts. More precisely, a system is divisible if there is a time evolution $\mathcal{T}_{\Delta t} : \mathcal{S} \to \mathcal{S}_1 \times \mathcal{S}_2$ that takes the state of a system and returns the state of two separate isolated systems. It is reducible if $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$, the state of the system is the state of the two subsystems. Most of the time these two concepts overlap, but they are separate as we can see in the following examples.

Consider a planarian flatworm. If we cut it into multiple pieces, each part will grow into a new worm. Yet, the state of each worm is not the state of multiple worms put together. The planarian is divisible into multiple worms but it is not reducible to multiple worms. Consider a magnet. Its state is described, to a first approximation, by the position of the north and south poles. Yet, we cannot divide the magnet into a north and south pole. The magnet is reducible to the poles but is not divisible into the poles.

Consider a muon. It will decay into an electron and two neutrinos. But the state of a muon is not the state of an electron and two neutrinos. The muon is divisible into the three particles but it is not reducible to them. Consider a proton. Its state is given by the state of its three valence quarks and the gluon field. Yet, we cannot divide the proton into the three isolated quarks. The proton is reducible to quarks, but not divisible into quarks.

The fundamental property of a quantum particle, as introduced here, is not indivisibility: it's irreducibility. It is the state that cannot be further partitioned, not the material itself.

Even more subtle is the difference between divisibility and decomposability, the idea that the object has parts. That is, a system is decomposable if there is a law $+ : \mathcal{S} \times \mathcal{S} \to \mathcal{S}$ to combine two states to get a third. While irreducibility says nothing on whether a quantum particle can be divided, it implicitly assumes that it can be decomposed. Irreducibility describes the relationship between the state of the whole and the state of the parts, therefore the existence of parts is implicit in the definition. But it also tells us that we are limited in what description we can give to these parts: we cannot isolate them as we want and act on them independently because if we were able to do that we could assign them an independent state.

Therefore, while we must be able to talk about parts of a quantum particle, not all questions will make sense. Can we ask what fraction of a quantum particle is in a particular region of space? Yes, since the particle can be spatially decomposed some amount of material may be inside the volume. Can we ask where that same fraction will be at a later time? No, since the particle is irreducible we cannot describe the evolution of a part.

Note that if the material is infinitesimally reducible (i.e. the classical case), reducibility/decomposability/divisibility are all possible. This is why one may not dwell on their difference and use the word "part" interchangeably in the three cases. Now that we assume irreducibility, it is better to use more precise language to avoid confusion. Therefore we will call *fragments* the infinitesimal parts of the material that cannot be described independently. We will call *components* the parts that the material can be decomposed into. A component will have a state associated with it, and therefore will always be a collection of fragments. As in the classical case, a component may be the difference between two states, therefore one should not imagine it always "contained" in the composed system.

The proper way to think of a quantized material is

therefore the following: a continuum of infinitesimal fragments clumped together to form quantum particles, the smallest amount of material to which we can assign a state. With this picture in mind, we can see, even before developing the math, how some features typical of quantum behaviors emerge.

The first feature is the importance of statistics. Consider a fragment of a quantum particle in an initial state. As we cannot track its motion, it can correspond to any fragment of the particle in its final state. That is, the motion of a fragment of a particle is a random walk within the evolution of a particle.

The second feature is non-locality. Suppose we have a quantum particle distributed in space. Suppose we exert a force only on a subregion of its distribution. As the particle is irreducible, we cannot interact with only part of it: we must also affect the part where no force is exerted. We cannot simply interact with the part at a specific region, as this would allow us to assign a state to that part.

Wave-particle duality is also accounted for. The quantum particle behaves like a wave in the sense that its parts form a distribution in space. Yet, it behaves as a single unit, as the state is irreducible and cannot be split.

The most striking feature, though, is that we can predict the functional form for particle decay over time. Suppose a quantum particle spontaneously decays into two or more particles. Such a process cannot be deterministic as the number of d.o.f. increases. As the decay is spontaneous, it is not triggered by the environment (i.e. external state). It must be triggered by the internal dynamics of the fragments. As we assumed the internal dynamics is unknown, we cannot learn anything about it by looking at the product of the decay. In particular, all states have to have equal chance of decay or we could tell that the internal dynamics is more stable for some states. The probability of decay is constant for all states at all times. The probability of survival follows the exponential distribution.

As we see qualitatively how the features typical of quantum mechanics naturally emerge from this picture, we are ready to make this more quantitatively and mathematically precise.

IX.B. Irreducible material and complex vector spaces

It is time to turn these concepts into a precise mathematical formulation as we did for classical mechanics. We want to formalize the state space for an *irreducible material* as a quantum particle will be a unit amount of such material.

As for the classical material, the irreducible material is thought to be decomposable: each amount can be seen as the sum of different components (e.g. spatially distributed). Therefore V.1 applies and the state space Ψ of an irreducible material is an abelian group.

Yet, the description of each component cannot be more accurate than the description of the whole. To capture this idea precisely we need to characterize the aspects of the unstated part that influence the state definition. How do we do that? We can look at the classical derivation for guidance. In that case, we looked at the set of transformations that increased/decreased the amount of material. The idea is that we can think of these transformations as characterizing what the state can and cannot say about the infinitesimal parts.

Suppose we have a state c_1 for an amount a_m of classical material. Suppose we assigned a numeric identifier to each particle of the material. This identifier would be just a dummy label we assign, not a physical quantity. As the classical material is homogeneous, permutations between classical particles within the same state have no effect. Therefore the ordering given by the identifier does not matter. Identifying a classical particle within a composite state is equivalent to picking a number from a random variable A uniformly distributed over an interval of size a_m.

Suppose we have another state $c_2 = \tau(a)c_1$ where the only difference is an increase/decrease of the amount of material by a factor a. The random variable B that identifies a particle of c_2 is now a uniform distribution $B = aA$ where the range is the new amount of material: there are now proportionally more/less labels to assign. This is also the only possible transformation that can link the two uniform random variables A and B. Therefore increase/decrease of material is the only operation that is captured by the state among those that affect the labeling of the fragments. In other words: the classical state does not care about the permutation of the fragments, but it cares how many fragments are there.

For the irreducible material we follow the same strategy and study the transformations that affect the labeling of what the irreducible state cannot tell us: the configuration of each fragment. Similarly, the irreducible state will not care about permutations of fragment configurations, but it will care how many fragment configurations are there.

Suppose we have a state ψ_1 for an amount a_m of irreducible material characterized by one d.o.f. The configurations of the infinitesimal fragments will be identified by two values, as a d.o.f. is two-dimensional. Each fragment configuration is identified by a pair of independent random variables (A, B), which are, as before, just dummy variables we use to name the fragment configurations. As the material is homogeneous, the distribution is uniform or we could tell something about the fragment configurations. We choose A and B to be uniform distributions of interval $\sqrt{a_m}$: identifying a fragment configuration is like picking a point at random over a square of area a_m, the amount of material.

Suppose we have another state ψ_2, where the labeling for the fragment configurations has changed. We can label the fragment configurations of ψ_2 in the same way, with two independent uniform distributions (C, D). But

since the labeling is the only thing that has changed, we must be able to relate the new distribution over (C, D) to the old distribution over (A, B). The relationship between the two pairs of variables must be linear since all distributions are uniform. $C = aA + bB$. To preserve the square distribution we must then have $D = -bA + aB$. We can write $\psi_2 = \tau(a, b)\psi_1$ as a and b are the only parameters of the transformation. Or more compactly, $\psi_2 = \tau(c)\psi_1$ where $c = a + \imath b$.

The transformation $\tau(c)$ increases/decreases the area by the modulus $|c|^2 = a^2 + b^2$, which physically corresponds to the increase/decrease of amount of material. It also rotates the square by the phase $\theta_c \equiv arg(c)$, which corresponds to a change in correlation: if $\theta_c = 0$, $C = |c|A$ are clearly correlated while if $\theta_c = \pi$, $C = -|c|A$ are clearly anti-correlated.

This line of reasoning is fairly abstract and it may seem to have no bearing with quantum states. Therefore, jumping a bit ahead, let's see if we can already find quantitative consequences of this model and see whether they are consistent with our knowledge of quantum mechanics.

If the amount of material is connected to the area of two uniform distributions, it should combine like the variance of a random variable. And in fact it does. Suppose X and Y are two random variables,

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\,\sigma_X \sigma_Y \rho_{X,Y} \tag{IX.1}$$

where σ^2 is the variance, σ the standard deviation, $\rho_{X,Y} = COV(X, Y)/\sigma_X \sigma_Y$ is the Pearson correlation coefficient(ref?). As $-1 \leq \rho_{X,Y} \leq +1$, the total variance can be greater or smaller than the sum of the variance of the components. Now suppose ψ is the state for a unit amount of irreducible material (i.e. a quantum particle). We can write $c\psi = (c_1 + c_2)\psi = c_1\psi + c_2\psi$. We have:

$$|c_1 + c_2|^2 = |c_1|^2 + |c_2|^2 + 2|c_1||c_2| \cos(\theta_{c_2} - \theta_{c_1}) \tag{IX.2}$$

The two expressions are formally identical and it shouldn't come as a surprise. The more material we have, the more labels for the fragment configurations to choose from, the bigger the spread of the random variables that assign the labels. $-1 \leq \cos(\theta_{c_2} - \theta_{c_1}) \leq +1$ takes the place of the Pearson correlation coefficient. Quantum interference can then be understood in terms of the combination of correlated/anti-correlated distributions.

If the phase determines the correlation between two states, then it should only be meaningful when comparing two states. And in fact only differences in phases are physically distinguishable in quantum mechanics. A more meaningful name for the phase, then, would be *correlation angle*.

If quantum states are uniform distributions over fragment configurations, we should see a hint of that when computing distributions over observables. And in fact we do. Suppose we have $Q\psi = q\psi$, an eigenstate of some observable. Then it is also an eigenstate of the infinitesimal transformation of the conjugate variable $1 + \frac{Qdp}{\imath\hbar}$. Since it has a symmetry over p, the distribution has to be equal at each value of p: the distribution is uniform.

If quantum particle states are really just unit amounts of material uniformly distributed, we should be able to calculate the Shannon entropy associated with them. Using our dummy variables, $I(\psi) = -\int \rho \ln(\rho) dAdB = -\int 1ln(1)dAdB = 0$. For a quantum mixed state $\tilde{\rho} = \sum \tilde{\rho}_i |\psi_i\rangle\langle\psi_i|$, we have $I(\tilde{\rho}) = -\sum \tilde{\rho}_i \ln \tilde{\rho}_i - \tilde{\rho}_i I(\psi_i) = -\sum \tilde{\rho}_i \ln \tilde{\rho}_i = -tr(\tilde{\rho} \ln \tilde{\rho})$. And in fact this is the Von Neumann entropy (ref?) used in quantum information, that implicitly assumes all states provide the same amount of entropy (i.e. zero).

If two quantum particle states ψ_1 and ψ_2 are really just two states at equilibrium with the same entropy, we should always be able to find a deterministic and reversible evolution to connect the two. And in fact we can. Take the plane identified by ψ_1 and ψ_2. Find the infinitesimal rotation $1 + \frac{Od\alpha}{\imath\hbar}$ along that plane. $\psi_2 = e^{\frac{O\alpha}{\imath\hbar}}\psi_1$ is the unitary evolution of the other. Note that this is *not* always possible in classical mechanics, as two states c_1 and c_2 may have different entropy.

In other words, our simple model is not only consistent with quantum mechanics but it has predictive power: it provides intuition that is well supported by these findings. It is not "just a mathematical trick".

If we increase the number of independent d.o.f. the group of transformations remains the same. While we have more fragment configurations to label, each d.o.f. is labeled independently. Therefore we have n uniform distributions, each over a square of the same amount of material. This means that, as we combine components, the labels have to combine in the same way along the different d.o.f.: the increase/decrease of material and the change in correlation has to be the same across all d.o.f.

Mathematically, the state space Ψ for an irreducible material is a complex vector space. It is so because it is the state space for a decomposable material (i.e. an abelian group) where we can increase/decrease the amount and change the correlation of the internal description (i.e. a set of transformations conveniently parametrized by a complex number).

Proposition IX.3. The state space Ψ for an irreducible material is a vector space over $\mathbb{C}$.

Justification. We claim Ψ is an abelian group. Ψ is the state space for a decomposable homogeneous material and is therefore an abelian group by V.1.

We claim the set of physically distinguishable transformations T that relabel the fragment configurations is isomorphic to $\mathbb{C}$. Consider an amount $a_m \in \mathbb{R}$ of irreducible material characterized by one d.o.f. The configuration of a fragment of the material is characterized by two random variables $A, B : \Omega \to \mathbb{R}$ where Ω is the sample space for the configuration of the fragments. As the material is homogeneous, A and B can be chosen such that the distribution ρ of material over such variables is uniform over a range of $\sqrt{a_m}$ for both variables. We have $a_m = \int \rho dA \wedge dB = \int dA \wedge dB$.

Consider a second amount $\hat{a}_m$ of irreducible material where we changed the labeling of the fragment configurations. Its fragment configurations can also be characterized by a uniform distribution over two variables $C, D : \Omega \to \mathbb{R}$ over a range of $\sqrt{\hat{a}_m}$. As the two amounts of material differ just by the labeling of fragment configuration, (C, D) is a function of (A, B). We have $\hat{a}_m = \int dC \wedge dD = \int |J| dA \wedge dB$ where $|J|$ is the Jacobian determinant. As both distributions are uniform, $|J|$ has to be constant. The transformation has to be linear: $C = aA + bB$ and $D = -bA + aB$. As the change in amount of material, the only physically distinguishable quantity associated with the fragment configurations, is fully specified by a and b, no other transformation is physically relevant for the state. Let $\tau : \mathbb{C} \to T$ represent such a transformation parametrized by $c = a + \imath b$. $|c|^2 = |J| = (a^2 + b^2) = \hat{a}_m / a_m$ is the ratio of the two amounts of materials and is physically distinguishable. Define on T an addition $+ : T \times T \to T$ and a multiplication $* : T \times T \to T$ such that $\tau(z_1) + \tau(z_2) = \tau(z_1 + z_2)$ and $\tau(z_1) * \tau(z_2) = \tau(z_1 * z_2)$, $z_1, z_2 \in \mathbb{C}$, the sum and product of the transformation is equal to the sum and product of their respective factors. The phase $arg(c)$ under addition is physically distinguishable as it can affect the modulus of the result. τ is an isomorphism between T and $\mathbb{C}$ as fields. Assume now that the material is characterized by multiple independent d.o.f. The transformation on each d.o.f. is defined independently by $c_i \in \mathbb{C}$. As $\hat{a}_m / a_m$ is the same under all d.o.f. $|c_i| = |c_j|$. The relationship has to always hold under addition as well, therefore $arg(c_i) = arg(c_j)$. $T \cong \mathbb{C}$ also for multiple d.o.f.

We claim Ψ is a vector space over $\mathbb{C}$. The abelian group Ψ can be extended with the operations defined by T as was done in V.2 for the infinitesimally reducible case. □

IX.C. Complex valued distributions

In the previous sections we saw how a finite amount of classical material was described by a distribution over the state variables of infinitesimal parts. For an irreducible material, we cannot assign states to the fragments, therefore we cannot properly talk about state variables. Therefore we call *fragment variables* the quantities that we use to label the fragment configurations, and it will be over these variables that we will define the distribution of irreducible material.

Naturally, we cannot give a complete description for the configuration of a fragment, therefore we cannot give a joint distribution for all fragment variables at the same time.[29] We can, though, find a maximal set of fragment variables: one that can be used to provide the best description possible. While there isn't a unique maximal set, we can show that one such set is provided by the fragment variables $q^i : \mathcal{Q} \to \mathbb{R}$ that define the units required to describe the system.

In fact, we cannot add a new d.o.f. (as this would mean changing $\mathcal{Q}$) and we cannot add a new fragment variable in an existing d.o.f. (as this would fully specify said d.o.f.). Therefore the set of fragment variables q^i already provides a maximal set of fragment variables. This means that a state ψ must tell us for each point in $\mathcal{Q}$ the amount and correlation angle of the material. That is: the state will be identified by $\psi : \mathcal{Q} \to \mathbb{C}$, the wave function over $\mathcal{Q}$.

The square modulus at each point will give us the density of the material. As in the classical case, such density must be integrable to give us the total amount. Therefore the function $\psi(q) \in L^2(\mathcal{Q}, \mu)$ is square integrable. The norm $|\psi|^2 = \int_{\mathcal{Q}} |\psi(q)|^2 dq$ will give us the amount of material associated with the state.

As the material is homogeneous, all its quantum particles are made of the same amount. We can set, by convention, the amount of material for a particle to be unitary, therefore quantum particle states are the subset of Ψ such that $|\psi|^2 = 1$. As we have seen, changing the phase by a constant only affects the unstated part. Therefore normalized distributions that differ by a total phase are not physically distinguishable and represent the same quantum particle state.

Proposition IX.4. The state space Ψ for an irreducible material is isomorphic to a subspace of the space of Lebesgue square integrable continuous functions. That is $\Psi \cong \Psi(\mathcal{Q}, dq^i) \subseteq C(\mathcal{Q}, \mathbb{C}) \cap L^2(\mathcal{Q}, \mu, \mathbb{C})$ under the isomorphism $\Upsilon_q : \Psi \leftrightarrow \Psi(\mathcal{Q}, dq^i)$ where $\Psi(\mathcal{Q}, dq^i)$ is the space of wave functions over q^i. The state space for a quantum particle is isomorphic to the projective space $\mathsf{P}(\Psi)$.

Justification. We claim the manifold $\mathcal{Q}$ that defines the units to describe a fragment of an irreducible material also provides a maximal set of fragment variables: no other fragment variable can be added. Let $\mathcal{Q}$ be the manifold that defines the units. Then for each state $\psi \in \Psi$ there exists a distribution $\rho_\psi(q^i)$ that represents the quantity of material at each point. Suppose this distribution could be further refined. Then another independent variable k would exist such that for each state $\psi \in \Psi$ we would have a distribution $\rho_\psi(q^i, k)$. k cannot be part of a new d.o.f. as it would use a unit defined independently from $\mathcal{Q}$. This is a contradiction as $\mathcal{Q}$ defines all units. k cannot lie in the same d.o.f. as of one of the q^i or it would provide a complete description for the fragment. This is a contradiction as the material is irreducible. $\mathcal{Q}$ provides a maximal set of fragment variables.

We claim $\Psi \cong \Psi(\mathcal{Q}, dp^i) \subseteq C(\mathcal{Q}, \mathbb{C}) \cap L^2(\mathcal{Q}, \mu, \mathbb{C})$ as a vector space. The maximal physical description of a fragment configuration is given by a point in $\mathcal{Q}$ parametrized by q^i and by a transformation $\tau \in T$ parametrized by $c \in \mathbb{C}$. For each $\psi \in \Psi$ there must exist a wave function $\psi(q^i) : \mathcal{Q} \to \mathbb{C}$ that provides the maximal physi-

[29] This recovers the notion of complementarity introduced by Bohr(ref?).

cal description for the state. $\rho_\psi(q^i) = |\psi(q^i)|^2$ as the square modulus of $\psi(q^i)$ is the increase/decrease from the unit amount of material. $\rho_\psi(q^i) \in C(\mathcal{Q}, \mathbb{R})$ by IV.4. $\rho_\psi(q^i) \in C^1(\mathcal{Q}, \mathbb{R})$ by the same arguments in the justification for V.4. Therefore $\psi(q^i) \in C(\mathcal{Q}, \mathbb{C})$. $\rho_\psi(q^i) \in L^1(\mathcal{Q}, \mu, \mathbb{R})$ by the same arguments in the justification for V.3. Therefore $\psi(q^i) \in L^2(\mathcal{Q}, \mu, \mathbb{C})$. Let $\Upsilon_{q^i} : \Psi \to C(\mathcal{Q}, \mathbb{C}) \cap L^2(\mathcal{Q}, \mu, \mathbb{C})$ be the function that given a state returns its wave function. As the wave function captures the maximal description of the system, two different states must represent different distributions. $\forall \psi_1, \psi_2 \in \Psi, \psi_1 \neq \psi_2 \implies \Upsilon_{q^i}(\psi_1) \neq \Upsilon_{q^i}(\psi_2)$. Υ_{q^i} is injective. Υ_{q^i} is a bijection between Ψ and $\Psi(\mathcal{Q}, dq^i) \equiv \Upsilon_{q^i}(\Psi)$. Let $\psi = \psi_1 + \psi_2$, then $\Upsilon_{q^i}(\psi) = \Upsilon_{q^i}(\psi_1) + \Upsilon_{q^i}(\psi_2)$ as combining the distributions means combining the statistical description of the fragment configurations. Let $\psi_1 = \tau(c)\psi_2$, then $\Upsilon_{q^i}(\psi_1) = c\Upsilon_{q^i}(\psi_2)$ as $\tau(c)$ transforms the fragment configuration description by the factor c. Υ_{q^i} is a homomorphism between vector spaces. Ψ is isomorphic to $\Psi(\mathcal{Q}, dq^i) \subseteq C(\mathcal{Q}, \mathbb{C}) \cap L^2(\mathcal{Q}, \mu, \mathbb{C})$.

We claim the state space for a quantum particle is isomorphic to the projective space $\mathsf{P}(\Psi)$. Let $\psi \in \Psi$ be the state of a quantum particle. $\tau(e^{i\theta})\psi$ with $\theta \in \mathbb{R}$ is not physically distinguishable from ψ as phases are only physically distinguishable through addition. $\tau(\sqrt{\rho})\psi$ with $\sqrt{\rho} \in \mathbb{R}^+ \backslash \{1\}$ cannot represent a quantum particle as it has a different amount of material. There exists only one distinguishable quantum particle state in the equivalence class $\psi \sim \tau(c)\psi$ with $c \in \mathbb{C}$. The state with no amount of material does not correspond to any particle state. The quotient space $\mathsf{P}(\Psi) \equiv (\Psi \backslash \{0\})/\sim$ is isomorphic to the state space for quantum particles. □

IX.D. Inner product and operators

In both the reducible and irreducible case, the material is distributed over $\mathcal{Q}$. The q^i do not label the states of a finite amount of material, but the configurations for the infinitesimal parts. For a classical material, the infinitesimal parts can be given a state, have their own topological space and a set of local state variables (q^i, p_i) to identify them. We can use the richness of differential geometry. For an irreducible material, the infinitesimal parts cannot be given a state, cannot be fully identified by a set of variables. We have to use other mathematical tools, typically vector spaces. As we will often use linear combinations of states, we will write, for brevity, $c\psi$ instead of $\tau(c)\psi$. That is: each complex number still represents a state transformation.

For the tool set to be on par with classical mechanics we need to address three requirements. We need to be able to compare two states and ask what fraction of them share the same configuration: this will give us an inner product. We need to be able to express distributions and expectations for different quantities: this will give us self-adjoint linear operators. We then need to identify the self-adjoint linear operators for the quantities k_i and p_i and to be able to express the distribution of material in terms of those. Let's start with the first.

The idea is that, given two states ψ_1 and ψ_2, we should be able to tell how much of ψ_1 is prepared in the same way as ψ_2. That is $\psi_1 = \alpha\psi_2 + \psi_3$ is the combination of the component that is prepared like ψ_2 (e.g. same distributions in position and momentum) and a component that is prepared completely differently. Let $\mathsf{P}_{\psi_2} : \Psi \to \Psi$ be the operator that gives us the component $\mathsf{P}_{\psi_2}(\psi_1) \equiv \alpha\psi_2$ prepared like ψ_2. This operator is linear in ψ_1: the component of the combined system that is prepared like ψ_2 is the sum of the subcomponents that are prepared like ψ_2. It also gives us the same result if applied once or twice: once we found the component that is prepared like ψ_2, it remains the same. P_{ψ_2} is therefore a projection: a linear operator such that $\mathsf{P}_{\psi_2}(\mathsf{P}_{\psi_2}(\psi_1)) = \mathsf{P}_{\psi_2}(\psi_1)$.

If ψ_1 is already prepared like ψ_2, then its projection on ψ_2 will be equal to ψ_1. Conversely, if the projection of ψ_1 is equal to itself, then it means that the state is already prepared like ψ_2. That is: $\mathsf{P}_{\psi_2}(\psi_1) = \psi_1 \Leftrightarrow \psi_1 = \alpha\psi_2$. Because of this, we can define a non degenerate complex product $\langle\cdot|\cdot\rangle : \Psi \times \Psi \to \mathbb{C}$ such that $\mathsf{P}_{\psi_2}(\psi_1) \equiv \frac{\langle\psi_2|\psi_1\rangle}{|\psi_2|^2}\psi_2$.

The projections have other useful properties. Because the material is irreducible, if some fraction of ψ_1 is prepared like ψ_2, then the same fraction of ψ_2 is prepared like ψ_1. If not, studying the fragments of one state would give us a better (or worse) description for the fragments of the other state. Therefore we have $\frac{|\mathsf{P}_{\psi_2}(\psi_1)|^2}{|\psi_1|^2} = \frac{|\mathsf{P}_{\psi_1}(\psi_2)|^2}{|\psi_2|^2}$. This tells us that $|\langle\psi_2|\psi_1\rangle|^2 = |\langle\psi_1|\psi_2\rangle|^2$.

Lastly, if we take the projection of ψ_1 along ψ_2 and we project it again on ψ_1, we end up with a component of ψ_1. As this is a part of ψ_1 prepared as ψ_1, its state is just ψ_1 rescaled, with no change in correlation. That is: $\mathsf{P}_{\psi_1}(\mathsf{P}_{\psi_2}(\psi_1)) = a\psi_1$ with $a \in \mathbb{R}^+$. This tells us that $\langle\psi_2|\psi_1\rangle = \langle\psi_1|\psi_2\rangle^\dagger$.

Because of the properties of the norm and the projections, the product is positive definite, linear and conjugate symmetric. In other words, the state space Ψ for an irreducible material is an inner product space derived by the norm and projections.

Proposition IX.5. The state space Ψ for an irreducible material is an inner product with $\langle\cdot|\cdot\rangle : \Psi \times \Psi \to \mathbb{C}$ such that $\langle\psi_1|\psi_2\rangle = \int_{\mathcal{Q}} \psi_1^\dagger(q^i)\psi_2(q^i)d^n q$.

Justification. We claim there exists a map $|\cdot|^2 : \Psi \to \mathbb{R}$ such that $|\psi|^2 \geq 0\ \forall\psi \in \Psi$ and $|\psi|^2 = 0 \Leftrightarrow \psi = 0$. Let $|\cdot|^2 : \Psi \to \mathbb{R}$ be the map that returns the amount of material for a given state. Let $\psi \in \Psi$ be a state. $|\psi|^2 = \int \rho(q^i)d^n q = \int \psi^\dagger(q^i)\psi(q^i)d^n q \geq 0$. $|\psi|^2 = 0 \Leftrightarrow \psi = 0$ as the only state with no material is the empty state.

We claim $\forall\psi \in \Psi$ there exists a projection $\mathsf{P}_\psi : \Psi \to \Psi$ with $\psi \in \Psi$. Let $\psi \in \Psi$ be a state. Let $\mathsf{P}_\psi : \Psi \to \Psi$ be an operator that returns the component of a given state that is prepared as ψ. $\mathsf{P}_\psi(c_1\phi_1 + c_2\phi_2) = c_1\mathsf{P}_\psi(\phi_1) + c_2\mathsf{P}_\psi(\phi_2)$ is linear as the component of the composition prepared like ψ is the composition of the components so prepared.

$\mathsf{P}_\psi \circ \mathsf{P}_\psi = \mathsf{P}_\psi$ is a projection because the result of P_ψ is prepared as ψ.

We claim $\mathsf{P}_\psi(\phi) = \phi \Leftrightarrow \phi = c\psi$ for some $c \in \mathbb{C}$. Let $\psi, \phi \in \Psi$ such that $\mathsf{P}_\psi(\phi) = \phi$. The whole state ϕ is prepared as ψ. We can obtain ϕ from ψ by a relabeling of the fragment configurations. $\phi = c\psi$ for some $c \in \mathbb{C}$. Let $\psi, \phi \in \Psi$ such that $\phi = c\psi$ for some $c \in \mathbb{C}$. Then all of ϕ is prepared as ψ. $\mathsf{P}_\psi(\phi) = \phi$. Therefore $\mathsf{P}_\psi(\phi) = \phi \Leftrightarrow \phi = c\psi$.

We claim $|\psi|^2|\mathsf{P}_\psi(\phi)|^2 = |\phi|^2|\mathsf{P}_\phi(\psi)|^2$. Let $\psi, \phi \in \Psi$. As the material is irreducible, ψ and ϕ cannot be used to describe smaller parts. The fraction of ϕ that is prepared as ψ must be equal to the fraction of ψ that is prepared as ϕ. $\frac{|\mathsf{P}_\psi(\phi)|^2}{|\phi|^2} = \frac{|\mathsf{P}_\phi(\psi)|^2}{|\psi|^2}$. Multiplying by $|\psi|^2|\phi|^2$ we have $|\psi|^2|\mathsf{P}_\psi(\phi)|^2 = |\phi|^2|\mathsf{P}_\phi(\psi)|^2$. This holds for the empty state as well. Let ϕ be the empty state. $|\phi|^2 = 0$ because it has no material and $|\mathsf{P}_\psi(\phi)|^2$ as no component of the empty state is prepared like ψ.

We claim $\mathsf{P}_\psi(\mathsf{P}_\phi(\psi)) = r\psi$ for some $r \in \mathbb{R}^+$. Let $\psi, \phi \in \Psi$. Consider $\mathsf{P}_\psi(\mathsf{P}_\phi(\psi))$: we first take the component of ψ prepared like ϕ and then take the component of that prepared like ψ. The overall operation is simply returning a part of ψ. This must correspond to the same exact configuration for the fragments. The only thing that can change is the amount of material. Therefore $\mathsf{P}_\psi(\mathsf{P}_\phi(\psi)) = r\psi$ for some $r \in \mathbb{R}^+$.

We claim that the state space Ψ for an irreducible material has an inner product $\langle\cdot|\cdot\rangle : \Psi \times \Psi \to \mathbb{C}$. $|\cdot|^2$ and P_ϕ satisfy the requirements in A.4. Ψ is an inner product space with $\langle\cdot|\cdot\rangle : \Psi \times \Psi \to \mathbb{C}$ such that $\langle\psi|\phi\rangle\psi = |\psi|^2\mathsf{P}_\psi(\phi)$.

We claim that the inner product can be expressed in terms of the wave functions as $\langle\psi_1|\psi_2\rangle = \int_{\mathcal{Q}} \psi_1^\dagger(q^i)\psi_2(q^i)d^nq$. The amount of material associated with a state $\psi \in \Psi$ is $\langle\psi|\psi\rangle = |\psi|^2 = \int_{\mathcal{Q}} \psi^\dagger(q^i)\psi(q^i)d^nq$. The inner product associated with the L^2 norm is $\int_{\mathcal{Q}} \psi_1^\dagger(q^i)\psi_2(q^i)d^nq$ with $\psi_1, \psi_2 \in \Psi$. $\langle\psi_1|\psi_2\rangle = \int_{\mathcal{Q}} \psi_1^\dagger(q^i)\psi_2(q^i)d^nq$.

□

Note that the inner product only has an indirect physical definition, based on the norm and the projections. These two concepts are therefore more useful to gain physical intuition as the first does not have a straightforward physical meaning.

Also note that the observations made after V.4 for the classical case are still valid. Ψ cannot be a complete metric space for the norm induced by the inner product as it does not include discontinuous functions. Therefore it is only a pre-Hilbert space. For mathematical convenience, one can take its completion with respect to the norm induced by the inner product and obtain a Hilbert space. Yet, such object no longer represents a physically meaningful set of states.

Therefore we need to make sure that the rest of our definitions and justifications do not require a complete metric space to be valid or they would not be physically meaningful. In particular, we don't want to use the usual link $\psi(q) = \langle q|\psi\rangle$ between the wave function and the inner product as the material cannot be prepared at a specific value when q is a continuous variable. Similarly, we cannot show that to each state variable is associated a self-adjoint operator $Q = \sum q|q\rangle\langle q|$ simply by expressing it in terms of eigenstates as they may not represent physical states.

We proceed in the following way. Suppose we have a fragment variable $f : \mathcal{Q} \to \mathbb{R}$. Consider the product $f(q)\psi(q)$: it returns the wave function weighted by the value of f. We can also think of f as an operator that takes a complex function of $\mathcal{Q}$ and returns another complex function of $\mathcal{Q}$ that represents the distribution of the expectation over the fragments. The operator is linear since $f(q)(a\psi_1(q) + b\psi_2(q)) = af(q)\psi_1(q) + bf(q)\psi_2(q)$. The operator is self-adjoint since $f^\dagger = f$. Consider $\int_{\mathcal{Q}} \psi^\dagger(q)f(q)\psi(q)dq = \int f(q)|\psi(q)|^2dq$. It integrates the value of f weighted by the amount of the material: it is the expectation of the fragment variable f over the distribution.

While $\mathcal{Q}$ is enough to define the system of units, it is not enough to define all the fragment variables over which we can define the distributions of irreducible material. The conjugate variables $p_i = \hbar k_i$ cannot be expressed as a function of q^i, which reiterates the idea that we cannot specify all variables for each fragment. Suppose there is another maximal set of fragment variables. Then those variables will form a topological space $\hat{\mathcal{Q}}$ such that each ψ can be expressed as a wave function $\psi(\hat{q}) : \hat{\mathcal{Q}} \to \mathbb{C}$.

As each state ψ can be identified by a wave function on either space, there must be an operator $\Upsilon^q_{\hat{q}}(\psi(q)) = \psi(\hat{q})$ that converts the wave function on one space to the other.[30] As a linear combination of states corresponds to a linear combination of wave functions, $\Upsilon^q_{\hat{q}}$ is a linear operator. As converting the space again must give us the original wave function, $\Upsilon^{\hat{q}}_q \circ \Upsilon^q_{\hat{q}} = I$ is unitary. We can now define a linear operator $F_{[\hat{q}]} \equiv f(\hat{q})$ corresponding to fragment variable $f : \hat{\mathcal{Q}} \to \mathbb{R}$. Using $\Upsilon^q_{\hat{q}}$ we can express $F_{[q]} = \Upsilon^{\hat{q}}_q \circ F_{[\hat{q}]} \circ \Upsilon^q_{\hat{q}}$ as an operator over complex functions of $\mathcal{Q}$. It will still be linear and self-adjoint.

That is: for each fragment variable f we have a corresponding linear self-adjoint operator $F_{[q^i]}$ that acts on wave functions over q^i, which can be expressed as function multiplication in a suitable choice of $\hat{\mathcal{Q}}$. The (normalized) expectation of f is given by $\frac{1}{|\psi|^2}\int \psi^\dagger(q^i)F_{[q^i]}(\psi(q^i))d^nq$.

Proposition IX.6. For each fragment variable f there exists a topological space $\hat{\mathcal{Q}}$ such that $f : \hat{\mathcal{Q}} \to \mathbb{R}$ and a vector space isomorphism $\Upsilon_{\hat{q}^i} : \Psi \leftrightarrow \Psi(\hat{\mathcal{Q}}, d\hat{q}^i)$ that takes a state $\psi \in \Psi$ and returns a wave function $\psi(\hat{q}^i)$ from $\Psi(\hat{\mathcal{Q}}, d\hat{q}^i)$ the space of wave functions

[30] Note that a neighborhood of q is not necessarily a neighborhood of $\hat{q}$. They are only if q and $\hat{q}$ can be expressed as a bijection of each other $\hat{q}^i(q^i)$: they are different coordinates of the same manifold.

over $\hat{\mathcal{Q}}$. There also exists a corresponding self-adjoint linear operator $F_{[q^i]} : \Psi(\mathcal{Q}, dq^i) \to C(\mathcal{Q}, \mathbb{C})$ such that $\int \psi^\dagger(q^i) F_{[q^i]}(\psi(q^i)) d^n q$ is the expectation of f multiplied by the amount of material.

Justification. We claim for each fragment variable f there exists a topological space $\hat{\mathcal{Q}}$ such that $\hat{q}^i : \hat{\mathcal{Q}} \to \mathbb{R}^n$ provides a maximal description and $f : \hat{\mathcal{Q}} \to \mathbb{R}$. Let f be a fragment variable. Let $\hat{q}^i$ be a set of fragment variables such that $f = f(\hat{q}^i)$. If $\hat{q}^i$ does not provide a maximal set of fragment variables then there exist other fragment variables that refine the description. Extend $\hat{q}^i$ with such variables. If still not maximal, continue extending. As the number of possibilities charted by $\hat{q}^i$ cannot exceed the number of possibilities charted by q^i, at some point we'll reach a maximal set. We call $\hat{\mathcal{Q}}$ the topological space charted by the fragment variables $\hat{q}^i$.

We claim there exists a vector space isomorphism $\Upsilon_{\hat{q}^i} : \Psi \leftrightarrow \Psi(\hat{\mathcal{Q}}, d\hat{q}^i)$. Let $\psi \in \Psi$ be a state. $\hat{q}^i$ provides a maximal set of fragment variables. There must exist a wave function $\psi(\hat{q}^i) : \mathbb{R}^n \to \mathbb{C}$ associated with ψ such that it gives the maximal description for the state. Let $\Upsilon_{\hat{q}^i} : \Psi \to \Psi(\hat{\mathcal{Q}}, d\hat{q}^i)$ be the map that associates a state with its wave function over $\hat{\mathcal{Q}}$. $\Upsilon_{\hat{q}^i}$ is a vector space isomorphism for the same reason discussed in IX.4

We claim that for each fragment variable $f : \hat{\mathcal{Q}} \to \mathbb{R}$ there exists a corresponding self-adjoint linear operator $F_{[\hat{q}^i]} : \Psi(\hat{\mathcal{Q}}, d\hat{q}^i) \to C(\hat{\mathcal{Q}}, \mathbb{C})$ such that $\int \psi^\dagger(\hat{q}^i) F_{[\hat{q}^i]}(\psi(\hat{q}^i)) d^n \hat{q}$ is, if it exists, the expectation of f multiplied by the amount of material. The expectation value of f multiplied by the amount of material is given by $\int f(\hat{q}^i)\rho(\hat{q}^i) d^n \hat{q}$ where $\hat{q}^i$ provide a maximal set of fragment variables such that $f = f(\hat{q}^i)$. We have $\int f(\hat{q}^i)\rho(\hat{q}^i) d^n \hat{q} = \int \psi^\dagger(\hat{q}^i) f(\hat{q}^i) \psi(\hat{q}^i) d^n \hat{q}$. Let $F_{[\hat{q}^i]} : \Psi(\hat{\mathcal{Q}}, d\hat{q}^i) \to C(\hat{\mathcal{Q}}, \mathbb{C})$ such that $F_{[\hat{q}^i]}(\psi(\hat{q}^i)) = f(\hat{q}^i)\psi(\hat{q}^i)$. $F_{[\hat{q}^i]}$ is a linear operator. $F_{[\hat{q}^i]}$ is self-adjoint as $\int (F_{[\hat{q}^i]}(\psi(\hat{q}^i)))^\dagger \psi(\hat{q}^i) d^n \hat{q} = \int (f(\hat{q}^i)\psi(\hat{q}^i))^\dagger \psi(\hat{q}^i) d^n \hat{q} = \int \psi^\dagger(\hat{q}^i) f(\hat{q}^i) \psi(\hat{q}^i) d^n \hat{q} = \int \psi^\dagger(\hat{q}^i) F_{[\hat{q}^i]}(\psi(\hat{q}^i)) d^n \hat{q}$.

We claim that for any fragment variable f there exists a corresponding self-adjoint linear operator $F_{[q^i]} : \Psi(\mathcal{Q}, dq^i) \to C(\mathcal{Q}, \mathbb{C})$ such that $\int \psi^\dagger(q^i) F_{[q^i]}(\psi(q^i)) d^n q$ is, if it exists, the expectation of f multiplied by the amount of material. Let f be a fragment variable. There exists a corresponding self-adjoint linear operator $F_{[\hat{q}^i]} : \Psi(\hat{\mathcal{Q}}, d\hat{q}^i) \to C(\hat{\mathcal{Q}}, \mathbb{C})$ where $\hat{\mathcal{Q}}$ provides a maximal set of fragment variables. As both $\hat{\mathcal{Q}}$ and $\mathcal{Q}$ provide a maximal set, we can define $\Upsilon^{q^i}_{\hat{q}^i} \equiv \Upsilon_{\hat{q}^i} \circ \Upsilon^{-1}_{q^i}$ which transforms a wave function over $\mathcal{Q}$ into a wave function over $\hat{\mathcal{Q}}$. $\Upsilon^{q^i}_{\hat{q}^i}$ is an isomorphism as $\Upsilon_{\hat{q}^i}$ and Υ_{q^i} are isomorphisms. $\Upsilon^{q^i}_{\hat{q}^i}$ is unitary as $\Upsilon^{\hat{q}^i}_{q^i} \circ \Upsilon^{q^i}_{\hat{q}^i} = (\Upsilon^{q^i}_{\hat{q}^i})^{-1} \circ \Upsilon^{q^i}_{\hat{q}^i} = I$. We can extend $\Upsilon^{q^i}_{\hat{q}^i}$ on all continuous functions and define $F_{[q^i]} \equiv \Upsilon^{\hat{q}^i}_{q^i} \circ F_{[\hat{q}^i]} \circ \Upsilon^{q^i}_{\hat{q}^i}$. $F_{[q^i]}$ is a self-adjoint linear operator. $\int \psi^\dagger(q^i) F_{[q^i]}(\psi(q^i)) d^n q = \int \psi^\dagger(\hat{q}^i) F_{[\hat{q}^i]}(\psi(\hat{q}^i)) d^n \hat{q}$ is, if it exists, the expectation of f multiplied by the amount of material. □

The last thing to do is characterize the conjugate variables k_i and p_i. To be precise, we want to find $K_{i[q^i]}$, the operators that act on the wave functions over q^i corresponding to fragment variables k_i. We know that the operators have to be covariant: under a change of units $\hat{q}^j = \hat{q}^j(q^i)$ we must have $\hat{K}_{j[q^i]} = \partial_{\hat{q}^j} q^i K_{i[q^i]}$. In other words: we are looking for the set of linear operators that obey the chain rule. This turns out to be the space of derivations, therefore $K_{i[q^i]}$ will be the derivatives in the direction of the corresponding q^i. They have to be self-adjoint, so they will be of the form $K_{i[q^i]} = a\imath\partial_i$ where a is a real number. We choose a to be negative by convention and to be unitary so that k_i is expressed in units of inverse q^i as in the classical case. Therefore we have $K_{i[q^i]} = -\imath\partial_i$ and $P_{i[q^i]} = -\imath\hbar\partial_i$.

If we assume $\mathcal{Q}$ to be $\mathbb{R}^n$, the functional $\Upsilon^{q^i}_{p_i} : \Psi(\mathcal{Q}, dq^i) \leftrightarrow \Psi(\mathcal{P}, dp_i)$ that converts a wave function in position to a wave function in conjugate momentum is the Fourier transform. As the fragment variables for any power of q^i and p_i must exist, the wave functions in both expressions are infinitely differentiable: the space of the wave functions $\Psi(\mathcal{Q}, dq^i) = S(\mathcal{Q}, q^i)$ is the Schwartz space of rapidly decreasing smooth functions.

Proposition IX.7. Let $\mathcal{Q} = \mathbb{R}^n$. The operator $P_{i[q^i]} : \Psi(\mathcal{Q}, dq^i) \to C(\mathcal{Q}, \mathbb{C})$ associated with the conjugate quantity of q^i is $P_{i[q^i]} = -\imath\hbar\partial_i$. The isomorphism $\Upsilon^{q^i}_{p_i} : \Psi(\mathcal{Q}, dq^i) \leftrightarrow \Psi(\mathcal{P}, dp_i)$ that changes the variable of the wave function from q^i to p_i is $\Upsilon^{q^i}_{p_i}(\psi(q^i)) = \frac{1}{(\sqrt{2\pi})^n} \int_{\mathcal{Q}} e^{\frac{q^i p_i}{\imath\hbar}} \psi(q^i) d^n q$. The space of wave functions $\Psi(\mathcal{Q}, dq^i) = S(\mathcal{Q}, q^i)$ where $S(\mathcal{Q}, q^i)$ is the Schwartz space of rapidly decreasing smooth functions.

Justification. We claim the operator associated with k_i is $K_{i[q^i]} = -\imath\partial_i$. Let $K_{i[q^i]} : \Psi(\mathcal{Q}, dq^i) \to C(\mathcal{Q}, dq^i)$ be the linear operator associated with k_i as expressed on the space of wave functions over q^i. Such an operator must be covariant under the transformation $\hat{q}^j = \hat{q}^j(q^i)$. That is: $\hat{K}_{j[q^i]} = \partial_{\hat{q}^j} q^i K_{i[q^i]}$. In particular, it needs to be covariant when changing only one variable: $\hat{q}^1 = \hat{q}^1(q^1)$, $\hat{q}^j = q^j$ where $j = 2...n$. By A.5 $K_{i[q^i]} = c\partial_i$ for some $c \in \mathbb{C}$. $K_{i[q^i]}$ is self-adjoint: $K_{i[q^i]} = \imath r\partial_i$ for some $r \in \mathbb{R}$ by A.6. k_i is expressed in units of inverse q^i therefore $|r| = 1$. By convention, we choose r to be negative. $K_{i[q^i]} = -\imath\partial_i$.

We claim the operator associated with p_i is $P_{i[q^i]} = -\imath\hbar\partial_i$. As $p_i = \hbar k_i$, their expectation is proportional by a factor of $\hbar$. $P_{i[q^i]} = \hbar K_{i[q^i]} = -\imath\hbar\partial_i$.

We claim the operator that changes the variable of the wave function from q^i to p_i is $\Upsilon^{q^i}_{p_i}(\psi(q^i)) = \frac{1}{(\sqrt{2\pi})^n} \int_{\mathcal{Q}} e^{\frac{q^i p_i}{\imath\hbar}} \psi(q^i) d^n q$. Let $\Upsilon^{q^i}_{p_i} : \Psi(\mathcal{Q}, dq^i) \to \Psi(\mathcal{P}, dp_i)$ be the operator that maps a wave function over $\mathcal{Q}$ into a wave function over $\mathcal{P}$, the space charted by p_i. $\Upsilon^{q^i}_{p_i}$ must exist and be unique (up to a total phase) as $\Upsilon^{q^i}_{p_i} = \Upsilon_{p_i} \circ \Upsilon^{q^i} = \Upsilon_{p_i}(\Upsilon_{q^i})^{-1}$ is an isomorphism.

$\Upsilon^{q^i}_{p_i}$ must be a linear transform that takes the operator $P_{i[q^i]} = -\imath\hbar\partial_i$ as expressed over q^i to $P_{i[p_i]} = p_i$ as expressed over p_i. Such transformation is the Fourier transform. $\Upsilon^{q^i}_{p_i}(\psi(q^i)) = \frac{1}{(\sqrt{2\pi})^n}\int_{\mathcal{Q}} e^{\frac{q^i p_i}{\imath\hbar}}\psi(q^i)d^n q$.

We claim $\Psi(\mathcal{Q}, dq^i) \subseteq S(\mathcal{Q}, q^i)$. Let $\psi(q^i) \in \Psi(\mathcal{Q}, dq^i)$. $(P_{i[q^i]})^m\psi(q^i) = (-\imath\hbar\partial_i)^m\psi(q^i)$ represents the mth power of the conjugate momentum operator applied to the given state. As the quantity is physically meaningful, such an operation must be defined for any m therefore $\psi(q^i) \in C^\infty(\mathcal{Q})$. Let $\psi(p_i) = \Upsilon^{q^i}_{p_i}(\psi(q^i))$ be the wave function over momentum space. $(Q^i_{[p_i]})^m\psi(p_i) = (+\imath\hbar\partial_{p_i})^m\psi(p_i)$ represents the mth power of the q^i operator applied to the given state. As the quantity is physically meaningful, such an operation must be defined for any m therefore $\psi(p_i) \in C^\infty(\mathcal{P})$. As $\psi(q^i)$ is the Fourier transform of an infinitely smooth function, it decreases faster than any inverse power of q^i. That is: $\forall m \in \mathbb{Z}^+ \exists c \in \mathbb{C}$ such that $|\psi(q^i)| < \frac{|c|}{|q^i|^m}\ \forall |q^i| > 1$. $\Psi(\mathcal{Q}, dq^i) \subseteq S(\mathcal{Q}, q^i)$.

We claim $\Psi(\mathcal{Q}, dq^i) = S(\mathcal{Q}, q^i)$. Let $\psi(q^i) \in \Psi(\mathcal{Q}, dq^i)$ and $\phi(q^i) \in S(\mathcal{Q}, q^i)$ such that $\psi(q^i) \neq \phi(q^i)$. There exists a choice of $a, b, j, k \in \mathbb{Z}^+$, $j, k \leq n$ such that $\|\psi(q^i) - \phi(q^i)\|_{j,a,k,b} \neq 0$ where $\|\psi(q^i)\|_{j,a,k,b} = sup|(q^j)^a(\partial_k)^b\psi(q^i)|$ as $S(\mathcal{Q}, q^i)$ is a Frechet space induced by that family of seminorms. The operator $(Q^j_{[q^i]})^a(P_{k[q^i]})^b$ applied to $\psi(q^i)$ and $\phi(q^i)$ will provide different distributions for the associated fragment variables. $\phi(q^i)$ is physically distinguishable and is associated with a physical state. $\phi(q^i) \in \Psi(\mathcal{Q}, dq^i)$. □

We now have all the basic features of the state space of a quantum particle. And we obtained them by rederiving them from first principles instead of using some formal analogy. The similarities with the classical framework are simply due to the similarities of the starting points.

It should be evident, though, how the correspondence between mathematical and physical objects is not as satisfying as in the classical case. For instance, while any differentiable function of $T^*\mathcal{Q}$ can be a state variable for a classical particle, not all self-adjoint linear operators can be associated with a fragment variable. Consider $f : \mathbb{R} \to \mathbb{R}$ for which $f(r) = -r$ if r is an integer and $f(r) = r$ otherwise. Clearly $f(q)$ does not preserve the topology and cannot be used as a fragment variable. Yet, we can still create a self-adjoint linear operator, with the same spectra as the one associated with q, just with the eigenvalues switched. The space of physically meaningful operators, then, should be further constrained. A reasonable requirement would be that they preserve the differentiable/integrable structure of the Schwartz space $S(\mathcal{Q}, q^i)$ as classical state variables preserved the differentiable structure of $T^*\mathcal{Q}$. Unfortunately, this is not as easy to justify.

The topology of $\mathcal{P}$, the space charted by conjugate momentum, depends globally on the topology of $\mathcal{Q}$. For example, if $\mathcal{Q}$ is an n-dimensional torus, there is no infinity and $\mathcal{P}$ has a discrete topology. That is why, in the end, we restricted ourselves to $\mathbb{R}^n$. Moreover, we characterized the space of functions not because we required them to fall off at infinity, but because we required the ability to express polynomials of q^i and p_i. It is not immediately clear how this generalizes over a manifold: further work is needed.

While the link between math and physics may not be as elegant and as general as one would like, we still managed the goal: we identified the states of an irreducible material and saw how these are related to the states in quantum mechanics.

IX.E. Irreducibility

Now that we have fully characterized what we mean by an irreducible material, we can stipulate the following:

Assumption IV (Irreducibility)**.** *The system under study is composed of an irreducible homogeneous material and as a whole undergoes deterministic and reversible evolution.*

Rationale. The idea is that each amount of material has a state, the physically distinguishable configuration of the distribution, and an unstated part, the physically indistinguishable configuration of the fragments of the material. The system is deterministic and reversible in the sense outlined in assumption I: given the state of the system at one time (i.e. the whole distribution) we can predict/reconstruct the future/past states. But we cannot describe the evolution of the unstated part (i.e. the motion of the fragments).

As we saw, the state space recovered under this assumption is the one used in quantum mechanics. In the following sections we will derive two types of state evolution: the deterministic and reversible one, corresponding to unitary evolution (i.e. Schroedinger's equation), and a non-deterministic one useful to describe measurement interactions, corresponding to the projection postulate. Therefore, as the infinitesimal reducibility assumption leads to classical mechanics, the irreducibility assumption leads to quantum mechanics.

As we saw in assumption II, infinitesimal reducibility had its problems. How about the irreducibility assumption? How do they compare?

The methodological problem assumes a different connotation. In order for the assumption to hold, we need to show that there are no processes at our disposal that can probe the fragments of a quantum particle. For example: an electron as a whole interacts with a photon as a whole. Note that this does not require showing that no such process exists: conceptually specific parts of the photon may be interacting with specific parts of the electron. But if those processes are not available to us, say because we can only manipulate whole photons and they always interact with a whole electron, they cannot be used to create a more detailed picture. So we simply have to enumerate

the processes available to us, which is, at least in principle, feasible. Yet, we can never rule out that someone in the future may discover a new process.

The issues arising from a lack of perfect isolation are addressed. Each degree of freedom comes with its own unstated part (i.e. the motion of the fragments within that d.o.f.), therefore there is no pretense of a perfect description under perfect isolation. Since we cannot describe the nature of the chaotic motion, we cannot say whether it is internally or externally driven.

The problem of physically meaningless interactions is also considerably lessened. Even if we have multiple particles each with their state evolving deterministically, we can still imagine their unstated parts as interacting with each other. Therefore each quantum particle would not essentially reside in its own separate universe.

In these ways, the irreducibility assumption fares much better than the infinitesimal reducibility assumption. Yet, it is still plagued with some of the same problems.

The problem of incompatibility of measurements with deterministic evolution at all levels still remains. The unstated part of multiple irreducible particles may be interacting, but it cannot be used to pass information from one subsystem to another without making itself distinguishable.

The inability to define time using a deterministic and reversible evolution is still a problem. As the system is homogeneous and irreducible, the unstated part cannot give us a measurable quantity that changes in time as that would provide a description of the unstated part.

There is one additional problem, though, not present in the classical case. Under the irreducibility assumption states are necessarily at equilibrium: they all provide the same level of description for the fragments, they carry the same information entropy. In fact if the internal dynamics were out of equilibrium, the material would not be homogeneous which would contradict the assumption. Deterministic and reversible evolution (which we'll see corresponds to unitary evolution) is necessarily quasi static: at each moment a state must be well defined, at each moment the system is in equilibrium, the system changes slowly. Therefore quantum states, as formulated here, cannot and should not be expected to provide a valid description during non-deterministic evolution.

In light of this, consider a muon decay: the initial state of the muon and the final states of the electron and neutrinos can certainly be well described by quantum particle states.[31] We have no reason to expect, though, that while the muon is decaying into the resulting particles the state is always in equilibrium and therefore well described mathematically by a vector in a complex inner product space. After all, the state of a single spin-one-half particle cannot determine the future state of three spin-one-half particles.

This notion may seem in contradiction to the current practice. For example, the S-matrix, one of the main devices for calculating scattering amplitudes and cross sections, is derived by assuming unitary evolution.(ref?) Doesn't it imply we are assuming that deterministic and reversible evolution is happening at every instant?

If we look more closely, we realize that the process that the S-matrix captures is a limit where initial states are at $t \to -\infty$ and final states are at $t \to +\infty$, one that takes an infinite amount of time. That is, to calculate predictions for a fast out of equilibrium process we are approximating it using slow quasi-static evolution. This is actually not so uncommon: it is a standard first approximation in thermodynamics. The question is why would it work here? As we'll see in more detail when talking about the projection postulate, it's because of the irreducibility assumption itself.

Suppose we are trying to calculate a quantity that does not depend on the particular motion of the unstated part. Then, by its very nature, its value will be the same under an equivalent quasi-static evolution, because this too does not depend on the particular motion of the fragments. The probability distribution of the out-states during scattering is one such quantity. As the final statistical distribution is determined only by the initial state, it is not dependent on the particulars of the unstated part. In these cases, the use of deterministic and reversible evolution as a "stand-in" to calculate final statistical distributions is justified. However, it should not be taken literally as a physical model of what actually happens. There is no justification for assuming this process to be quasi-static deterministic and reversible evolution of quantum particle states.

In other words, the assumption of irreducibility already contains the seeds of its own demise. On one hand, it tells us that there is a component of chaotic motion within the system. On the other hand, it can only characterize the equilibrium dynamics, which clearly can't be the full range. Yet it also offers a way out: since the motion of the fragments is not accessible, we can predict statistical distributions that are independent of it.

In light of what we discussed, the irreducibility assumption is less flawed than infinitesimal reducibility but it still cannot be taken as fundamental, in the sense that we cannot take it to strictly apply to all systems. We should not expect to solve open problems, such as the arrow of time and the measurement problem, under the assumption as formulated in this work. □

IX.F. Quantum mechanics

We are now ready to write the equations of motion for an irreducible material. Deterministic and reversible evolution for the whole distribution necessarily means quasi-static evolution: there is a well defined state at

[31] In this context, a state is in equilibrium regardless of its being in a local minimum, easily overcome by fluctuations, or a global minimum, that persists indefinitely.

each moment in time. Moreover the components must evolve independently as we cannot learn more about one component by observing its evolution in different combinations: the evolution of the composition is the composition of the evolution. In other words: the evolution must preserve the inner product.

Note that formally there is a strong analogy with the reducible material case. In classical mechanics the state space structure was captured by a distribution over a symplectic manifold. Deterministic and reversible evolution was a transformation that preserved that structure, a symplectomorphism. In the quantum case the state space structure is captured by an inner product space. Deterministic and reversible evolution is a transformation that preserves that structure, a unitary transformation. That is: once we have defined what mathematical structure captures the definition of a state, deterministic and reversible evolution must necessarily preserve such structure.

Proposition IX.8. A deterministic and reversible evolution map $\mathcal{T}_{\Delta t} : \Psi \to \Psi$ for an irreducible material is a unitary operator. That is: $\langle \mathcal{T}_{\Delta t}\psi_1, \mathcal{T}_{\Delta t}\psi_2\rangle = \langle \psi_1, \psi_2\rangle$ where $\psi_1, \psi_2 \in \Psi$.

Justification. We claim $\mathcal{T}_{\Delta t} : \Psi \to \Psi$ is a unitary operator. Let $\psi_1, \psi_2 \in \Psi$ be two states. $\mathcal{T}_{\Delta t}(a\psi_1 + b\psi_2) = a\mathcal{T}_{\Delta t}\psi_1 + b\mathcal{T}_{\Delta t}\psi_2$ with $a, b \in \mathbb{C}$ as the composition of the evolution is the evolution of the composition: $\mathcal{T}_{\Delta t}$ is a linear operator. $\mathcal{T}_{\Delta t}\mathsf{P}_{\psi_1}(\psi_2) = \mathsf{P}_{\mathcal{T}_{\Delta t}\psi_1}(\mathcal{T}_{\Delta t}\psi_2)$ as the evolved part of ψ_2 that was prepared like ψ_1 is the part of the evolved ψ_2 that ends prepared like the evolved ψ_1. We have $\mathcal{T}_{\Delta t}\mathsf{P}_{\psi_1}(\psi_2) = \mathcal{T}_{\Delta t}\langle\psi_1, \psi_2\rangle\psi_1 = \langle\psi_1, \psi_2\rangle\mathcal{T}_{\Delta t}\psi_1 = \mathsf{P}_{\mathcal{T}_{\Delta t}\psi_1}(\mathcal{T}_{\Delta t}\psi_2) = \langle\mathcal{T}_{\Delta t}\psi_1, \mathcal{T}_{\Delta t}\psi_2\rangle\mathcal{T}_{\Delta t}\psi_1$. $\langle\psi_1, \psi_2\rangle = \langle\mathcal{T}_{\Delta t}\psi_1, \mathcal{T}_{\Delta t}\psi_2\rangle$. $\mathcal{T}_{\Delta t}$ is unitary. □

Proposition IX.9. A continuous deterministic and reversible evolution for an irreducible material admits a Hamiltonian operator $H : \Psi \to \Psi$ that allows us to write the laws of evolution as

$$\imath\hbar\partial_t\psi = H\psi$$

Justification. We claim the state evolves according to Schroedinger's equation. Let $\mathcal{T}_{dt} : \Psi \to \Psi$ be the evolution for an infinitesimal time interval dt. As it is unitary, it can be expressed as $\mathcal{T}_{dt} = I + \frac{Hdt}{\imath\hbar}$ where $H : \psi \to \psi$ is a self-adjoint operator. Let $\psi_t \in \Psi$ be a state and $\psi_{t+dt} \in \Psi$ its evolution after an infinitesimal time interval. We have $\mathcal{T}_{dt}\psi_t = \psi_{t+dt} = \psi_t + \frac{Hdt}{\imath\hbar}\psi_t$. $\imath\hbar\frac{\psi_{t+dt}-\psi_t}{dt} = \imath\hbar\partial_t\psi_t = H\psi_t$. □

We recognize the Schroedinger equation, the time evolution equation for a quantum particle state.

Note, though, that nothing tells us that the system must be in an eigenfunction of H or that the ground state is somehow preferred. In fact, the eigenfunctions of H may not be part of $S(\mathcal{Q}, q^i)$ and therefore may not be physical. In other words: we have no basis for the time independent Schroedinger equation. This is actually consistent with our assumptions. For example, for a system to reach its ground state it has to be able to radiate energy: it cannot be isolated. Therefore this dynamics cannot be described under the deterministic and reversible assumption. This already hints that deterministic and reversible evolution is not the only dynamics possible. Other considerations, outside of the scope of this work, can be used to recover this behavior (e.g. thermodynamics, decoherence, ...).

Also note that H operates on Ψ which, in the case of $\mathcal{Q} = \mathbb{R}^n$, corresponds to the space $S(\mathcal{Q}, q^i)$ of Schwartz functions. H therefore preserves infinite smoothness and integrability. This, again, reminds us that the space of all self-adjoint operators is far too vast, and does not map well to physically meaningful quantities.

IX.G. Time dependence and kinematic assumption

The next step would be to integrate the quantum description with time dependent evolution (i.e. relativistic mechanics) and the kinematic assumption (i.e. Lagrangian mechanics). We will only provide a sketch, though, without going through the details. Given that we are limited to single particles (i.e. no fields) without spin (i.e. minimal material only described by position), our results are limited to finding the Klein-Gordon equation modified with electromagnetic interactions. While this is still of note, the equation is of limited physical use and therefore may not warrant the space necessary for a more rigorous derivation.

As in the classical case, the first thing to do is to extend $\mathcal{Q}$ to include time. This gives us the same $\mathcal{M}$ as before. A quantum state is a complex valued distribution $\psi : \mathcal{M}_{t=t_0} \to \mathbb{C}$ over a hypersurface at constant time. An evolution $\widetilde{\psi} : \mathcal{M} \to \mathbb{C}$ is a complex valued function defined on the whole $\mathcal{M}$. The operator for the quantity conjugate to time is $E = \imath\hbar\partial_t$.

Next we need to extend integration. In the quantum case, this means to make sure the inner product $\langle\psi|\phi\rangle$ remains unchanged under all variable changes, including time. This allows us to describe the amounts of material and their distribution over fragments in a way that is independent of our description.

As in the classical case, we use s as the evolution parameter. Deterministic and reversible evolution $\mathcal{T}_{ds}$ will be a unitary transformation defined over this extended inner product, as the structure of the state space is conserved. Continuous evolution will admit an invariant Hamiltonian $\mathcal{H}$ such that $\imath\hbar\partial_s\psi = \mathcal{H}\psi$. As for the composite state evolution in the classical case, $\widetilde{\psi}$ remains unchanged by the evolution as it transports the distribution along s. Therefore $\mathcal{H}\widetilde{\psi} = \hbar\widetilde{\psi}$ is an eigenfunction of $\mathcal{H}$. As in the classical case, we use $\mathcal{H} = 0$ to identify the state space: the temporal d.o.f. is not independent and therefore there is a constraint among the variables. Therefore we have $\mathcal{H}\psi = 0$. This is the time dependent evolution equation. It means that the wave function does not change along s, the parameter of the infinitesimal

transformation generated by $\mathcal{H}$.

Note that states are indeed eigenfunctions of the invariant Hamiltonian $\mathcal{H}$: here we do have a justification, as opposed to the standard Hamiltonian H case. For example, a massive particle may have fragments with different energy, but they all must carry the same fraction of mass. Also note that $\mathcal{H}\psi = 0$ is basically the form of all equations (i.e. Klein-Gordon, Dirac, Maxwell, ...) in quantum field theory. The respective invariant Hamiltonian $\mathcal{H}$ is the operator associated with each equation.

Having handled the time dependent case, we can turn our attention to the kinematic equivalence assumption. While the trajectory of a fragment is not well defined in either phase space or physical space, what we can describe remains the same under the assumption. Therefore the distributions still need to be transported from state space to kinematic variables up to a constant factor. This means we still have a linear relationship between velocity and conjugate momentum. In terms of the quantum operators: $m\, d_s Q^\alpha = g^{\alpha\beta} P_\beta - A^\alpha = m\frac{[Q^\alpha, \mathcal{H}]}{\imath\hbar}$.

This leads to the same form for the invariant Hamiltonian $\mathcal{H} = \frac{1}{2m}((P_\alpha - A_\alpha)g^{\alpha\beta}(P_\beta - A_\beta) + m^2c^2) = \frac{1}{2m}((-\imath\hbar\partial_\alpha - A_\alpha)g^{\alpha\beta}(-\imath\hbar\partial_\beta - A_\beta) + m^2c^2)$. In the free particle case, we have $\frac{1}{2m}(-\hbar^2\partial^\alpha\partial_\alpha + m^2c^2)\widetilde{\psi} = 0$ which is the Klein-Gordon equation. We can also write the operator for kinetic momentum as $mU^\alpha = P^\alpha - A^\alpha = -i\hbar D^\alpha$ where $D^\alpha = \partial^\alpha + \frac{A^\alpha}{\imath\hbar}$ is the gauge covariant derivative. We have $(\frac{1}{2}mU^\alpha U_\alpha + \frac{1}{2}mc^2)\widetilde{\psi} = \frac{1}{2m}(-\hbar^2 D^\alpha D_\alpha + m^2c^2)\widetilde{\psi} = 0$ which is the Klein-Gordon equation with electromagnetic interaction.

Note how the gauge covariant derivative is, up to a constant, kinetic momentum. This motivates why it is such an important operator. As for the Klein-Gordon equation, the only thing it does is to impose that the norm of the four-velocity is the speed of light, a condition much more trivial than one may have anticipated.

It should be clear, even without the details, that all the assumptions come together even in the quantum case. They give more insight to all the pieces in a comprehensive way and help tear down the artificial walls between the different physical theories. This allows us to form a picture of fundamental physics that is more unified, something that has been sorely missing.

IX.H. Non-deterministic evolution and the projection postulate

As we made it clear from the start, deterministic and reversible evolution is an assumption. And as we have seen in the rationale for assumptions II and IV we cannot have determinism at all levels of decomposition while transferring information from one system to another.

In this section we will see how non-deterministic processes become necessary when describing measurements for irreducible systems. But we'll see that, under the same assumption of irreducibility, we are able to make a connection between the output of deterministic and non-deterministic processes.

Suppose we have a quantum particle state ψ we want to identify. Suppose that we have a detector that is able to capture the particle. If it's captured, we will know that the particle was in the region of space enclosed by the detector.

Suppose that we create an array of such detectors. This will allow us to detect the position of the particle with a resolution given by the aperture of each detector.

Suppose now, though, that the quantum particle state is spread over a region significantly greater than a single detector but smaller than the array. What will happen?

It is clear that it has to be absorbed as it is within the aperture of the whole array. It is also clear that it must be absorbed whole: as it is an irreducible system, the detector cannot partially interact with it, it cannot half absorb it. So a single detector will capture it. It is also clear that it cannot always be the same detector: that corresponds to states where the particle is spread within the aperture of that detector. In other words: as the particle hits the array, one detector triggers non-deterministically based on the distribution of the particle.

The process is non-deterministic in the sense that the state of the particle is not enough to predict which detector will trigger. Maybe it depends on the state or unstated part of the detector, maybe it depends on the unstated part of the particle or maybe on the environment. The point is that a single measurement does not tell us much about the state of the incoming particle.

We then repeat the process, making sure the incoming particle is prepared in the same way. Given a single "pure" incoming particle state, we get a "mixed" statistical distribution describing the output of the array.[32] The question is: how is the statistical distribution coming out of the detector related to the material distribution coming in with the particle?

If the detector does not introduce biases, then the final distribution is only a function of the incoming particle: all unbiased detectors will give the same result. The unstated part also cannot affect the final distribution of outcomes: it is not physically distinguishable. The final statistical distribution of the detector must match the incoming material distribution of the particle. That is if N is the total number of takes and N_A is the number of times the detector with aperture in the A region captured the particle, the ratio $\frac{N_A}{N}$ will be approximately the amount of material

[32] Note that this is different from the model used often to describe the quantum measurement problem, where the incoming system becomes entangled with the measuring device. We do not make any claim on what the final state of the particle should be: in many cases, like in a photomultiplier tube, the particle is gone and there is no state. We simply have one incoming state for the particle, prepared over and over exactly the same, and the outgoing state for the measurement device, which is a classical statistical distribution over the different measurement takes.

within the aperture $\int_A \psi^\dagger(q^i)\psi(q^i)d^nq$. If we assume an infinite number of takes, we expect them to coincide: $\int_A \psi^\dagger(q^i)\psi(q^i)d^nq = \lim_{N\to\infty} \frac{N_A}{N}$. But this is also the frequentist definition of probability. Therefore we have $\int_A \psi^\dagger(q^i)\psi(q^i)d^nq = P(\text{measuring particle in } A)$ thus recovering the probabilistic interpretation of the wave function.

We can conceptually extend this picture to the generic case. Suppose we know the outcomes of our measurement are parametrized by a fragment variable $\hat{q}$. We will have a linear functional $\Lambda^{\frac{1}{2}}_{\hat{q}=\hat{q}_0} : \Psi \to \mathbb{C}$ such that $|\Lambda^{\frac{1}{2}}_{\hat{q}=\hat{q}_0}|^2$ returns the amount of material prepared with the particular value $\hat{q}_0$. We can re-express this quantity in terms of the dual vector $\hat{q}_0$ corresponding to the linear functional $\Lambda^{\frac{1}{2}}_{\hat{q}=\hat{q}_0}$ and the inner product extended to all Lebesgue integrable functions: $|\Lambda^{\frac{1}{2}}_{\hat{q}=\hat{q}_0}|^2 = |\langle\hat{q}_0|\psi\rangle|^2$. This is also the probability that the outcome associated with $\hat{q}_0$ is selected by the measuring device, thus recovering the statistical nature of the quantum predictions.

Note that while this process is of a non-deterministic nature, it is of a specific non-deterministic nature. It requires that no other state can influence the final distribution as the match between the physical material distribution and the statistical measurement distribution was recovered only under the assumption that the detector does not introduce any bias. That is: the non-deterministic nature cannot be understood as an averaging over some unknown external state.

Also note that while we have a way to predict the distribution over final outcomes from the initial state, we have no way of describing what happens in between. When we assume that only the initial state matters, we are saying that we can disregard the particulars of the non-deterministic evolution. So much so that we can substitute it with a deterministic one to compute the probability distribution. In particular, given only the initial state, we have no way of predicting which set of final outcomes the system will project to as that depends on the actual choice of detector setup, nor when that projection actually happens.

While this may seem like an incomplete description, it is as complete as the irreducibility assumption allows. The state can be known and determines any final statistical distribution. The unstated part cannot be known, and is what gives the statistical distribution in the first place. If we were able to say something more about the evolution of the unstated part, we would be able to further describe it, which we cannot do without violating the assumption. If we cannot learn anything about the unstated part from the final distribution, then we are consistent with our assumption and we are able to use deterministic evolution and projection to determine the final statistical distribution. In other words: the inability to describe the unstated part not only constrains the state space and deterministic evolution, but also the set of non-deterministic processes by excluding those for which the final statistical distribution would depend on the unstated part itself.

We therefore have arrived at all the major underpinnings of quantum mechanics and see how they stem from assumption IV of irreducibility. This gives us a more satisfactory account of what quantum mechanics describes, although it does not claim to solve many of the open problems. It does, however, give good reasons as to why we have those problems and how the desire to know "what is really going on" is hindered by the very assumption that the quantum description is based on.

X. DISCUSSION AND FURTHER WORK

As we went through many details in the derivation, it is time to step back and conclude with a few general remarks.

The first remarkable aspect of this work is that something along these lines can be done. Regardless of whether all the minor details are addressed in the best possible way, the idea that so much can be derived from so little was not something we expected. In hindsight, the idea that physics should be founded on physical principles is nothing new: it just hasn't been thoroughly applied to basic mechanics. If this work had been done incrementally from the start, these concepts would have been developed gradually and the results would feel more mundane.

The second striking aspect is a shift in perspective that this work seems to suggest. The usual reductionist view is that you start with the rules of the small parts and then build up the composite whole. This work flips that perspective: it's not what the parts can tell you about the whole, it's what the whole can tell you about the parts.

The state space for classical particles is the one that allows us to express invariant distributions: that is a requirement that comes from the whole. The fragments of quantum particles do not even have a well defined description: the whole tells us nothing about their configuration. In retrospect, this should not be surprising. Nobody is going to hand us down the laws of infinitesimal parts: we induce what we can by manipulating finite systems. Therefore what the whole allows us to infer about its parts is physically fundamental.

The third notable aspect is that the assumptions have more to do with our ability to create a reliable description rather than essential features of the system itself. A system is not deterministic per se: it is a particular choice for its description (i.e. state) that may undergo deterministic evolution in some cases. In the same vein, a material is not homogeneous per se. Or infinitesimally reducible. In this context, the state, its properties and the laws of evolution are the description of the system under our idealization, not some intrinsic feature of the system.

The fourth intriguing aspect is that, when framing

classical and quantum mechanics in this way, there is a progression. First we assume a homogeneous infinitesimally reducible material. Then we assume a homogeneous irreducible material. Extrapolating from this progression, the next assumption would be an inhomogeneous irreducible material: where the material is not in constant equilibrium and its particles do not all look the same.

This different starting point may have potential connections to some of the problems we mentioned in the rationales for the assumptions and to other interesting physics questions in general. In the foreseeable future, however, we will work on the more mundane task of extending this work to field theories, for which a more concrete path is available.

XI. CONCLUSION

We have seen that much of fundamental physics can be derived from few simple physical assumptions. In doing so we obtained a more unified picture of the different branches of fundamental physics centered around the idea of deterministic and reversible evolution of distributions of homogeneous materials. Though the work is extensive we can summarize a few important conclusions with the following points:

- A state space is not just a collection of states for the system under study but it must also capture physical properties such as physical distinguishability (represented by a topology), the ability to count states (represented by a measure, symplectic form or metric tensor) or decomposability (represented by a vector space structure).
- Deterministic and reversible evolution is a one-to-one map that preserves the nature of the system (e.g. physical distinguishability, state count and decomposability). Therefore it is an isomorphism on the mathematical structure defined by the state space (e.g. homeomorphism, symplectomorphism, unitary transformation), not just a state-to-state bijection (i.e. isomorphism of sets).
- Classical Hamiltonian mechanics describes the evolution of a homogeneous material whose state is equivalent to the states of all its infinitesimal parts, the classical particles. The state space for classical particles is a symplectic manifold $(T^*\mathcal{Q}, \omega)$ as this allows us to define state-variable-invariant density distributions; deterministic and reversible evolution for such a state space is a symplectomorphism (i.e. a canonical transformation).
- Assuming classical particle states are fully characterized by their trajectory in space, the states also evolve according to Lagrangian mechanics and the motion is that of a massive particle under scalar/vector potential forces.
- Relativistic motion arises by properly handling time dependent deterministic and reversible evolution for the material (i.e. distributions over phase space) without further assumptions.
- Quantum (Hamiltonian) mechanics describes the evolution of a homogeneous material for which the state can only be reduced up to finite parts, the quantum particles, whose internal motion is undetermined. The state space for quantum particles is a complex inner product space $(\Psi, \langle\cdot,\cdot\rangle)$; deterministic and reversible evolution for such a state space is a unitary transformation.

While we are sure that some mathematical details can be improved, we feel that the overall "conceptual castle" is solid. The starting physical assumptions are simple, the translation into math is often straightforward and the intuitive ideas that emerge help unveil connections within and across different branches of math and physics. It is the tight fitting of these different pieces into a broader picture that validates the merit of the approach.

While the scope of this work is limited to refounding and reunderstanding physical theories already experimentally established, it would not be unexpected for a better grasp of old theories to lead to insights and ideas for new ones.

ACKNOWLEDGMENTS

We'd like to thank Erik Curiel, Jiahua Gu, Josh Hunt, Isaac Mooney and Pavel Okun for providing feedback on drafts of this manuscript, Mark Greenfield for further insights in the topological aspects and Alan Phillips for contributing figures. We additionally thank Paul Stankus for interest and support in the earlier stages of this project. Funding for this work was provided in part by the MCubed program of the University of Michigan.

APPENDIX A. MATHEMATICAL PROOFS

In this section we include mathematical demonstrations in support of the physical justifications of the paper.

Theorem A.1 (Extended Riesz theorem)**.** *Let (S, τ) be a locally compact Hausdorff space. Let $\Lambda = \{\Lambda_U : C(S) \to \mathbb{R}\}_{U\subseteq S}$ a family of positive linear functionals such that $\forall U \subseteq S$ $\Lambda_U = \Lambda_{\mathrm{int}(U)}$ and $\Lambda_{U_1} + \Lambda_{U_2} = \Lambda_{U_1\cup U_2} + \Lambda_{U_1\cap U_2}$ $\forall U_1, U_2 \subseteq S$. Then there exists a unique Borel measure μ such that $\Lambda_U(\rho) = \int_U \rho d\mu$ $\forall \rho \in C(S)$.*

Proof. We claim that at each $s \in S$ there exists a compact neighborhood $U \subseteq S$ endowed with a Borel measure μ_U such that $\Lambda_U(\rho) = \int_U \rho d\mu_U$ $\forall \rho \in C(S)$. S is locally compact. $\forall s \in S$ $\exists U \subseteq S$ such that U is compact. U with the subspace topology is a compact Hausdorff topological space on which is defined a positive linear functional

$\Lambda_U : C(S) \to \mathbb{R}$. For the Riesz representation theorem for linear functionals, there exists a unique Borel measure μ_U such that $\Lambda_U(c) = \int_U \rho d\mu_U \ \forall \rho \in C(S)$.

We claim that S is endowed with a unique Borel measure μ such that $\Lambda_U(\rho) = \int_U \rho d\mu$. Let $U \subseteq V \subseteq \mathcal{S}$ two compact subsets. Let $\rho \in C(S)$. $\Lambda_V(\rho) = \int_V \rho d\mu_V = \int_U \rho d\mu_V + \int_{V\setminus U} \rho d\mu_V$. Also $\Lambda_V(\rho) = \Lambda_U(\rho) + \Lambda_{V\setminus U}(\rho) = \int_U \rho d\mu_U + \int_{V\setminus U} \rho d\mu_{\overline{V\setminus U}}$. Therefore $\int_U \rho_c d\mu_U = \int_U \rho_c d\mu_V$. $d\mu$ is unique and does not depend on the choice of neighborhood. S is locally compact. It admits a cover $\{U_\alpha\}_{\alpha\in A}$ where each U_α is compact.

$$\begin{aligned}\Lambda_{\mathcal{S}}(c) &= \sum_{\alpha\in A} \Lambda_{U_\alpha}(c) - \sum_{\alpha\in A}\sum_{\beta\in A}^{\beta\neq\alpha} \Lambda_{U_\alpha\cap U_\beta}(c) \\ &= \sum_{\alpha\in A}\int_{U_\alpha} \rho_c d\mu - \sum_{\alpha\in A}\sum_{\beta\in A}^{\beta\neq\alpha}\int_{U_\alpha\cap U_\beta}(c)\rho_c d\mu \\ &= \int_{\mathcal{S}} \rho_c d\mu\end{aligned}$$

The measure is unique on the whole space. □

Theorem A.2. *Let $\mathcal{L}(S,\mu) \equiv \{\rho : S \to \mathbb{R} \mid \ |\int_U \rho d\mu| < \infty \ \forall U \subseteq S\}$. $\mathcal{L}(S,\mu) = L^1(S,\mu)$ where $L^1(S,\mu) = \{\rho : S \to \mathbb{R} \mid \ \int_S |\rho| d\mu < \infty\}$*

Proof. We claim $\mathcal{L}(S,\mu) \supseteq L^1(S,\mu)$. Let $\rho \in L^1(S,\mu)$. Let $U \subseteq S$. $\|\rho\|_U \equiv \int_U |\rho| d\mu < \int_S |\rho| d\mu < \infty$. $\|\rho\|_U = \|\rho^+\|_U + \|\rho^-\|_U$. $\|\rho^+\|_U < \infty$ and $\|\rho^-\|_U < \infty$. $|\int_U \rho d\mu| = |\int_U \rho^+ d\mu - \int_U \rho^- d\mu| = |\int_U \rho^+ d\mu| - |\int_U \rho^- d\mu| = \|\rho^+\|_U - \|\rho^-\|_U < \infty$. $\rho \in \mathcal{L}(S,\mu)$.

We claim $L^1(S,\mu) \supseteq \mathcal{L}(S,\mu)$. Let $\rho \in \mathcal{L}(S,\mu)$. Let $S^+ \equiv \{s \in S \mid \rho(s) > 0\}$. Let $S^- \equiv \{s \in S \mid \rho(s) < 0\}$. $|\int_{S^+} \rho d\mu| = |\int_S \rho^+ d\mu| = \|\rho^+\| < \infty$ and $|\int_{S^-} \rho d\mu| = |\int_S \rho^- d\mu| = \|\rho^-\| < \infty$. $\|\rho\| = \|\rho^+\| + \|\rho^-\| < \infty$. $\rho \in L^1(S,\mu)$.

We claim $L^1(S,\mu) = \mathcal{L}(S,\mu)$. $\mathcal{L}(S,\mu) \supseteq L^1(S,\mu)$ and $L^1(S,\mu) \supseteq \mathcal{L}(S,\mu)$. □

Theorem A.3. *Let (M,ω) a symplectic manifold. Let $f : (M,\omega) \to (M,\omega)$ an infinitesimal self-symplectomorphism. Let $S \in TM \mid f(\xi^a(m)) = \xi^a(m) + S^a(m)dt \ \forall m \in M$ be the infinitesimal displacement. There exists a function $H : M \to \mathbb{R}$ such that $S^a\omega_{ab} = \partial_b H$ and $H \in C^2(M,\mathbb{R})$.*

Proof. We claim the vector field $S \in T\mathcal{Q}$ admits a potential H such that $S^a\omega_{ab} = \partial_b H$. Let $v, w \in T_mM$ be two vectors defined at a point $m \in M$. Let v^a, w^b be their components. Let $v' \equiv fv, w' \in T_{f(m)}M$ be the pushforward of v, w by f. Since f is a symplectomorphism we have $v^a\omega_{ab}w^b = v'^a\omega_{ab}w'^b$. The vector components change according to $v'^a = \partial_b\xi(t+dt)^a v^b = (\delta^a_b + \partial_b S^a dt)v^b$. We have:

$$\begin{aligned}v^a\omega_{ab}w^b &= v'^a\omega_{ab}w'^b \\ &= (v^a + \partial_c S^a v^c dt)\omega_{ab}(w^b + \partial_d S^b w^d dt) \\ &= v^a\omega_{ab}w^b + (\partial_c S^a v^c \omega_{ab} w^b + v^a\omega_{ab}\partial_d S^b w^d)dt \\ &+ O(dt^2)\end{aligned}$$

$v^c w^b \partial_c S_b - v^a w^d \partial_d S_a = 0$ where $S_b \equiv S^a\omega_{ab}$. The relationship must be true for and pair of vector, therefore $\partial_a S_b - \partial_b S_a = curl(S_a) = 0$. S is a curl free vector field and admits a potential H such that $S_a = \partial_a H$.

We claim $H \in C^2(M,\mathbb{R})$. S is differentiable as it is the displacement of a symplectomorphism. The derivatives of H are differentiable as $S_a = \partial_a H$. H is twice differentiable. □

Theorem A.4. *Let V be a complex vector space. Let $|\cdot|^2 : V \to \mathbb{R}$ such that $|x|^2 \geq 0 \ \forall x \in V$ and $|x|^2 = 0 \Leftrightarrow x = 0$. Let $x \in V$ and $\mathsf{P}_x : V \to V$ be a family of projections (i.e. P_x linear and $\mathsf{P}_x \circ \mathsf{P}_x = \mathsf{P}_x$) such that: $\mathsf{P}_x(y) = y \Leftrightarrow y = ax$ for some $a \in \mathbb{C}$; $|x|^2|\mathsf{P}_x(y)|^2 = |y|^2|\mathsf{P}_y(x)|^2$; $\mathsf{P}_x(\mathsf{P}_y(x)) = rx$ for some $r \in \mathbb{R}$. Then V is an inner product space with the product $\langle\cdot,\cdot\rangle : V\times V \to \mathbb{C}$ defined by $\langle x,y\rangle x = |x|^2\mathsf{P}_x(y)$.*

Proof. We claim $\langle\cdot,\cdot\rangle$ is well defined. Let $x, y \in V$. $\mathsf{P}_x(\mathsf{P}_x(y)) = \mathsf{P}_x(y)$ as P_x is a projection. $\mathsf{P}_x(y) = ax$ for some $a \in \mathbb{C}$ by the first property of P_x. $\langle x,y\rangle = |x|^2 a$ exists and is unique.

We claim $\langle\cdot,\cdot\rangle$ is positive definite. Let $x \in V$. $\langle x,x\rangle x = |x|^2\mathsf{P}_x(x) = |x|^2 x$. Therefore $\langle x,x\rangle = |x|^2 \geq 0$ and $\langle x,x\rangle = |x|^2 = 0 \Leftrightarrow x = 0$.

We claim $\langle\cdot,\cdot\rangle$ is linear in the second argument. Let $x, y, z \in V$ and $a, b \in \mathbb{C}$. $\langle x, ay+bz\rangle x = |x|^2\mathsf{P}_x(ay+bz) = a|x|^2\mathsf{P}_x(y) + b|x|^2\mathsf{P}_x(z) = a\langle x,y\rangle x + b\langle x,z\rangle x$ by linearity of P_x. Therefore $\langle x, ay+bz\rangle = a\langle x,y\rangle + b\langle x,z\rangle$.

We claim $\langle\cdot,\cdot\rangle$ is conjugate symmetric. Let $x, y \in V$. $||x|^2\mathsf{P}_x(y)|^2 = |x|^4|\mathsf{P}_x(y)|^2 = |\langle x,y\rangle|^2|x|^2$. Using the second property of the projection we have $|\langle x,y\rangle|^2 = |x|^2|\mathsf{P}_x(y)|^2 = |y|^2|\mathsf{P}_y(x)|^2 = |\langle y,x\rangle|^2$. The modulus of the product is symmetric as a function of the arguments. Now consider $|x|^2|y|^2\mathsf{P}_x(\mathsf{P}_y(x)) = |x|^2\langle y,x\rangle\mathsf{P}_x(y) = \langle y,x\rangle\langle x,y\rangle x$. The product $\langle x,y\rangle\langle y,x\rangle \in \mathbb{R}$ by the third property of the projection. $\arg(\langle x,y\rangle) + \arg(\langle y,x\rangle)$ is 0 or π. π is excluded since $\langle x,x\rangle = |x|^2 \geq 0$. $\langle y,x\rangle$ and $\langle x,y\rangle$ have equal modulus and opposite phase. $\langle x,y\rangle = \langle y,x\rangle^\dagger$ □

Theorem A.5. *Let $C^1(\mathcal{Q},\mathbb{C})$ be the space of differentiable complex functions defined over $\mathcal{Q}$. Let $q : \mathcal{Q} \to \mathbb{R}$ a differentiable function. Let $K_q : C^1(\mathcal{Q},\mathbb{C}) \to C^1(\mathcal{Q},\mathbb{C})$ be a covariant linear operator, that is given an invertible differentiable function $\hat{q} : \mathbb{R} \to \mathbb{R}$ $K_{\hat{q}} = \partial_{\hat{q}} q K_q$. Then $K_q = a\partial_q$ for some $a \in \mathbb{C}$.*

Proof. We claim $K_q(f(q)) = \partial_q f K_q(q)$. Let $I : \mathbb{R} \to \mathbb{R}$ be the identify function, that is $I \circ f = f \ \forall f : \mathbb{R} \to \mathbb{R}$. $K_q(q) = K_q(I(q)) = \partial_q I(q) K_I(I) = K_I(I)$ by covariance. Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable invertible

function. $K_q(f(q)) = K_q(I(f(q))) = \partial_q f K_f(I(f)) = \partial_q f \partial_f I(f) K_I(I) = \partial_q f K_q(q)$. Now let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function non necessarily invertible. It can be written as $f = f_1 + f_2$ with f_1, f_2 differentiable and invertible. By linearity $K_q(f) = K_q(f_1) + K_q(f_2) = \partial_q f_1 K_q(q) + \partial_q f_2 K_q(q) = \partial_q f K_q(q)$.

We claim K_q is a derivation. Let f, g, h be differentiable functions such that $h(q) = f(q)g(q)$. $K_q(ln(h(q))) = \partial_q ln(h(q)) K_q(q) = \frac{1}{h}\partial_q h K_q(q) = \frac{1}{h} K_q(h(q)) = K_q(ln(f(q)g(q))) = K_q(ln(f(q))) + K_q(ln(g(q))) = \frac{1}{f} K_q(f(q)) + \frac{1}{g} K_q(g(q))$. Multiplying by $h = fg$ we have $K_q(h(q)) = gK_q(f(q)) + fK_q(g(q))$. K_q is a linear operator that satisfies the product rule. K_q is a derivation by definition.

We claim $K_q = a\partial_q$ for some $a \in \mathbb{C}$. K_q is a derivation. As $K_q(f(q)) = \partial_q f K_q(q)$ it is a derivation along q, which can be expressed as $K_q(q) = a\partial_q$ for some $a \in \mathbb{C}$. □

Theorem A.6. *Let $G \equiv C^1(\mathcal{Q}, \mathbb{C}) \cap L^1(\mathcal{Q}, \mathbb{C})$ be the space of differentiable and integrable complex functions defined over $\mathcal{Q}$. Let K be a linear adjoint operator of the form $a\partial_q$ with $a \in \mathbb{C}$. Let K be self-adjoint with respect to the inner product $\langle f, g\rangle = \int_{\mathcal{Q}} f^\dagger(q) g(q) dq$. Then $K = \imath r \partial_q$ with $r \in \mathbb{R}$.*

Proof. We claim $K = \imath r \partial_q$ with $r \in \mathbb{R}$. Let $f, g \in G$. We have $\langle f, Kg\rangle = \int_{\mathcal{Q}} f^\dagger a \partial_q g dq = f^\dagger g|_{\partial\mathcal{Q}} - \int_{\mathcal{Q}} a\partial_q f^\dagger g dq = \int_{\mathcal{Q}} (-a\partial_q f^\dagger) g dq = \int_{\mathcal{Q}} (-a^\dagger \partial_q f)^\dagger g dq = \langle -a^\dagger \partial_q f, g\rangle$ using integration by part and integrability of f and g. $\langle -a^\dagger \partial_q f, g\rangle = \langle Kf, g\rangle$ as K is self-adjoint. $-a^\dagger \partial_q = a\partial_q$. $a = -a^\dagger$. $a = \imath r$ with $r \in \mathbb{R}$. □

[1] Thomas, R. E., Rosa, A. J.: The analysis and design of linear circuits. Wiley (2001)

[2] Barron, R. F.: 8.2 Lumped-parameter analysis. In: Barron, R. F.: Industrial Noise Control and Acoustics. CRC Press (2002)

[3] Borutzky, W.: Bond graph methodology: development and analysis of multidisciplinary dynamic system models. Springer (2010)

[4] Freidin, R., Lasnik, H.: Syntax: critical concepts in linguistics. Routledge (2006)

[5] Friedman, D. P., Wand, M., Haynes, C.T.: Essentials of Programming Languages. MIT Press (2001)

[6] Lee, J. M.: Introduction to Topological Manifolds. Springer (2000)

[7] Lee, J. M.: Introduction to Smooth Manifolds. Springer (2003)

[8] Lang, S. L.: Algebra. Springer (2002)

[9] Young, N.: An Introduction to Hilbert Space. Cambridge Univ. Press, Cambridge, UK (1988)

[10] José, J. V., Saletan, E. J.: Classical Dynamics. Cambridge University Press, (1998)

[11] Shannon, C. E.: A mathematical theory of communication. The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, (1948)

[12] Jaynes, E. T.: Information theory and statistical mechanics. Statistical Physics 3, pp. 181–218, (1963)

[13] Gromov, M. L.: Pseudo holomorphic curves in symplectic manifolds. Inventiones Mathematicae 82, pp. 307–347, (1985)

[14] de Gosson, M. A.: The symplectic camel and the uncertainty principle: the tip of an iceberg?. Foundations of Physics 39, pp. 194–214, (2009)

[15] Stewart, I.: The symplectic camel, Nature 329, pp. 17–18, (1987)

[16] Treumann, R. A., Nakamura, R., Baumjohann, W.: Relativistic transformation of phase-space distributions. Annales Geophysicae (2011)

[17] Gourgoulhon, E.: 3+1 Formalism and Bases of Numerical Relativity. arXiv. http://arxiv.org/abs/gr-qc/0703035 (2007)

[18] Synge, J. L.: Classical dynamics. Encyclopedia of Physics Vol 3/1, Springer (1960)

[19] Lanczos, C.: The variational principles of mechanics, University of Toronto Press, (1949)

[20] Struckmeier, J.: Hamiltonian dynamics on the symplectic extended phase space for autonomous and non-autonomous systems. J. Phys. A: Math. Gen. 38, pp. 1257–1278, (2005)

[21] Einstein, A., Podolsky, B., Rosen, N.: Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. 47 777 (1935)