

# 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: October 14, 2016)

**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, 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, 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, studying trajectories is equivalent to studying state evolution, will give us Lagrangian mechanics together and limit the form of the Hamiltonian/Lagrangian to the one with scalar and vector potential forces.

We work strives as much as possible to be philosophically consistent, physically meaningful and mathematically precise.

### I. STATUS

Section III - Done.

Section IV - Done.

Section V - Done.

Section VI - Mostly done. Needs review.

Section VII - Mostly done. Needs review.

Section VIII - Preliminary work

Section IX - Preliminary work

Before or after each proposition there is a **discussion** that conceptually explains the motivation and the result in a less formal, more approachable way.

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

### II. ORGANIZATION AND STYLE

The work is organized into:

**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 why not) such an assumption makes sense (in a particular case).

**Propositions:** these characterize the mathematical properties of the physical concepts we are studying. A **justification** follows each proposition to show why such characterization follows from the physical properties of the system. While the style of a justification is formal, it is not a pure mathematical proof as it necessarily has to mix physical and mathematical arguments.

This should allow readers with different focus and background to concentrate on the parts that are more keen on.

### 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.

---

\* carcassi@umich.edu

### 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.<sup>1</sup>

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 indepen-

<sup>1</sup> 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 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?

dent 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 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  $9.81 \text{ m/s}^2$ ; if I want the anvil to reach the ground at  $x \text{ m/s}$  I have to drop it from  $y \text{ m}$ . 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 thermodynamics and systems theory, is often defined as *the set of variables needed to determine the future evolution of the system*. 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. black-body 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 will 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.  $\square$

## 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 Hausdorff and second countable. 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  $+1/2$ .

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.<sup>2</sup> Note that different outcomes for the same process can overlap. For example,  $4.12 \pm 0.05$  cm and  $4.13 \pm 0.05$  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,  $4.12 \pm 0.0005$  cm and  $4.13 \pm 0.0005$  cm do not overlap anymore.

We can also conceptually combine two different processes into a single one. 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.<sup>3</sup> 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. 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. The fact that the potential outcomes of a process are countable requires the space to be second countable.

**Proposition IV.1.** The state space  $\mathcal{S}$  of a physical system is a Hausdorff and second countable 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  $\mathcal{T}$ . 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  $\mathcal{T}$  be the set of all sets associated with all physical outcomes.  $\emptyset \in \mathcal{T}$  and is associated with impossible outcomes.  $\mathcal{S} \in \mathcal{T}$  and is associated with unavoidable outcomes. Let  $V_1, V_2 \in \mathcal{T}$ . 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  $P_2$ . This always

<sup>2</sup> The information provided by the process as measured by Shannon’s entropy is finite.

<sup>3</sup> The quantum case is similar: we measure marginal distributions and rule out states that are incompatible with them.

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  $\mathcal{T}$  is closed under intersection. Let  $V_1, V_2 \in \mathcal{T}$ . 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  $V_1, V_2$  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  $\mathcal{T}$  is closed under union.  $\mathcal{T}$  is a topology by definition.

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

We claim that  $\mathcal{S}$  is second countable. Consider the set of outcomes of a process. As each is identified by a finite number of digits, the set of outcomes is countable. This remains true when combining processes: the number of processes we can combine is finite as is the number of times we can prepare the same state and the set of intersections obtained from the countable outcomes of a finite number of processes is still countable. Consider a basis for the topology. Suppose it is of cardinality greater than countable. Then the basis would be able to distinguish among states in a way that no physical process ever could. The topology could distinguish among states that are not physically distinguishable. The basis for the topology cannot be of cardinality greater than countable. By contradiction,  $\mathcal{S}$  must be second countable.  $\square$

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 and second countable topological space.

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

With our state space defined, deterministic and reversible evolution corresponds to a bijective map between initial and final states. Moreover, a physical process that can distinguish final states is also a physical process that can distinguish initial states. That is: the topology is mapped and preserved by the deterministic and reversible evolution (i.e. initial and final states are

equally distinguishable). 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  $\mathcal{T}_{\Delta t} : \mathcal{S} \rightarrow \mathcal{S}$ .

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

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  $\mathcal{T}_{-\Delta t} : \mathcal{S} \rightarrow \mathcal{S}$  that returns the initial state given the final state.  $\mathcal{T}_{-\Delta t} \circ \mathcal{T}_{\Delta t} = \mathcal{T}_{\Delta t} \circ \mathcal{T}_{-\Delta t} = id$  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.  $\square$

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  $f : \mathcal{S}_1 \rightarrow \mathcal{S}_2$  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.  $\square$

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.<sup>4</sup> This is what the math is telling us, that if we said before that outcomes can't distinguish isolated points (i.e. the topology is second countable) then neither can maps. It is physical consistency that limits us to continuous functions.

<sup>4</sup> 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.

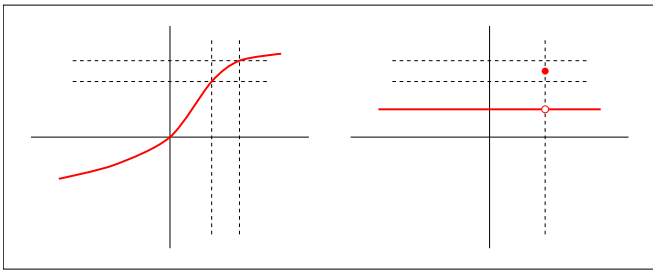


FIG. 1. A continuous function always maps finite precision knowledge in the vertical axis (i.e. an open set) to finite precision knowledge in the horizontal axis. A discontinuous function may map to finite precision (i.e. a closed set). If finite precision identification is ruled out (i.e. use of standard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

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 in 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 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. 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  $\bigcup_{1 \leq i \leq n} \mathbb{R}^{m_i}$ ).<sup>5</sup>

<sup>5</sup> TODO: Is there a more physically meaningful requirement? Could we have a set of states that are not locally described by a set of quantities?

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 requirements 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).<sup>6</sup>

**Proposition IV.5.** A *state variable* is a continuous map  $q : U \rightarrow \mathbb{L}$  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.  $\square$

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 cartesian 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  $q_1 : U \rightarrow \mathbb{L}_1$  and  $q_2 : U \rightarrow \mathbb{L}_2$  are said *independent* if  $\exists s$  such that  $q_1(s) = l_1, q_2(s) = l_2 \forall l_1 \in q_1(U), l_2 \in q_2(U)$ .

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

<sup>6</sup> 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, if there exists a one to one map between velocity and conjugate momentum they can both be physically well defined state variables.

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

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

#### 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  $q^i(s)$  to the final value  $q^i(\hat{s})$ . 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  $\hat{q}^i(\hat{s}) = q^i(s)$ , which we can always do as the evolution is deterministic and reversible. The level sets of  $\hat{q}^i$  are the evolved level sets 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.<sup>7</sup>

**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  $q^i \circ \mathcal{T}_{-\Delta t}$ .

To show the usefulness of transported state variables, we can prove the following:

**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 cardinality of the given set.

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

<sup>7</sup> 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 using different state variables or different values.

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  $\Delta q^1$  and  $\Delta q^2$  are ranges of possibilities for two independent variables, the total number of possibilities is  $\#(\Delta q^1)\#(\Delta q^2)$ . By construction, the set  $\Delta q^i$  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 trickier, and in fact will be part of the challenge in the following sections. 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. The issue is that while deterministic evolution will map points to points, not all point to point maps can be considered deterministic and reversible. Discrete sets already come with a way to measure and compare the number of elements which is preserved by any bijective map. With continuous sets, we need to do extra work to properly define a measure (or metric) that allows one to count states and possibilities.

## 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, 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 system made of infinitesimal homogeneous parts. What we'll find is that under the additional assumption that the system is infinitesimally reducible (i.e. its state is equivalent to the states of its 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 real 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 parts. 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.<sup>8</sup> 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.<sup>9</sup> All combined, this gives us the structure of an abelian group.

**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  $+: \mathcal{C} \times \mathcal{C} \rightarrow \mathcal{C}$  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 parts. There exists a unique zero element  $c + 0 = c$  and it represents the physically empty state

<sup>8</sup> 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 property is maintained by time evolution.

<sup>9</sup> 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.



(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 and second countable topological manifold as per IV.2. There exists an inverse  $- : \mathcal{C} \rightarrow \mathcal{C}$  such that  $c + (-c) = 0 \forall c \in \mathcal{C}$ .  $\mathcal{C}$  is an additive group by definition.  $\square$

### 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).<sup>10</sup>

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 decomposition. 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  $\rho_c : \mathcal{S} \rightarrow \mathbb{R}$  that returns the amount of material in the composite state  $c$  that is prepared in the particle state  $s$ . 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  $\rho_c$  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,  $\rho_c$  is a continuous function as discussed in IV.4.

As we combine the states of different parts, we sum the distributions over particle states  $\rho_c(s) = \rho_{c_1}(s) + \rho_{c_2}(s)$ : 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  $\rho_{c_1}(s) = a\rho_{c_2}(s)$ . 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  $\mathcal{C} \cong G \subseteq C(\mathcal{S}) \equiv \{\rho : \mathcal{S} \rightarrow \mathbb{R} \mid \rho \text{ is continuous}\}$ , 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 system 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<sup>11</sup> isomorphic to  $\mathbb{R}$ . Consider  $\tau : \mathbb{R} \rightarrow T$  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  $+: T \times T \rightarrow T$  and a multiplication  $*: T \times T \rightarrow T$  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.  $\tau$  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 each element  $\tau \in T$  is a map  $\tau : \mathcal{C} \rightarrow \mathcal{C}$ . The map has the following properties:  $(\tau_1 + \tau_2)c = \tau_1c + \tau_2c \forall \tau_1, \tau_2 \in T$  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  $\mathbb{R}$ .  $\forall c \in \mathcal{C} \exists ! \rho_c : \mathcal{S} \rightarrow \mathbb{R}$  returning the amount of material for each state  $s \in \mathcal{S}$ .  $\rho_c \in C(\mathcal{S})$  by

<sup>10</sup> We use this definition as it allows us to retroactively talk on somewhat equal grounds about classical point-particles and quantum particles, and underscore how physically (and mathematically) they are very similar yet very different objects.

<sup>11</sup> 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).

IV.4. Let  $\varrho : \mathcal{C} \rightarrow C(\mathcal{S})$  such that  $\varrho(c) \mapsto \rho_c$ . As the system is reducible, two distinct composite states must represent different distributions.  $\forall c_1, c_2 \in \mathcal{C}, c_1 \neq c_2 \implies \varrho(c_1) \neq \varrho(c_2)$ .  $\varrho$  is injective.  $\varrho$  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  $a$ .  $\varrho$  is a homomorphism as a vector space.  $\mathcal{C}$  is isomorphic to  $\varrho(\mathcal{C}) \subseteq C(\mathcal{S})$ .  $\square$

### V.C. Integration and measure

When the topology of  $\mathcal{S}$  is not discrete, the distribution  $\rho_c$  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  $\Lambda_U : \mathcal{C} \rightarrow \mathbb{R}$ .

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:  $\Lambda_U(a_1c_1 + a_2c_2) = a_1\Lambda_U(c_1) + a_2\Lambda_U(c_2)$ . 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  $\rho_c$  is positive everywhere, then the total amount is also positive. This makes  $\Lambda_U$  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<sup>12</sup>  $\mu$  for the size of sets 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, which gives  $\mathcal{S}$  the structure of a measure space, with a suitable Borel  $\sigma$ -algebra and measure  $\mu$ .

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

*Justification.* We claim there exists a positive linear functional  $\Lambda_U : \mathcal{C} \rightarrow \mathbb{R}$  for each  $U \subseteq \mathcal{S}$  such that  $\Lambda_U = \Lambda_{\text{int}(U)}$  where  $\text{int}(U)$  is the interior of  $U$ . Let  $\Lambda_U : \mathcal{C} \rightarrow \mathbb{R}$  be the functional that returns the amount of material in  $c$  compatible with the outcome associated with the open set  $U \subseteq \mathcal{S}$ .  $\Lambda_U$  is well-defined:  $U$  is associated with a physically distinguishable outcome.  $\Lambda_U$  is linear:  $\Lambda_U(a_1c_1 + a_2c_2) = a_1\Lambda_U(c_1) + a_2\Lambda_U(c_2)$  as it has to be consistent with the operations of composing states and increasing/decreasing the amount of material by a factor.  $\Lambda_U$  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  $\Lambda_U$  as  $\Lambda_{\text{int}(U)}$ .

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.  $|\Lambda_U(c)| = |\Lambda_{\text{int}(U)}(c)| < \infty$ .

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  $U_1 \cap U_2 = \emptyset$ ,  $\Lambda_{U_1 \cup U_2} = \Lambda_{U_1} + \Lambda_{U_2}$  as the material found in  $U_1 \cup U_2$  must be either in  $U_1$  or  $U_2$ . Suppose  $U_1 \cap U_2 \neq \emptyset$ ,  $\Lambda_{U_1 \cup U_2} = \Lambda_{U_1} + \Lambda_{U_2} - \Lambda_{U_1 \cap U_2}$  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.  $\Lambda_{U_1} + \Lambda_{U_2} = \Lambda_{\text{int}(U_1)} + \Lambda_{\text{int}(U_2)} = \Lambda_{\text{int}(U_1) \cup \text{int}(U_2)} + \Lambda_{\text{int}(U_1) \cap \text{int}(U_2)} = \Lambda_{U_1 \cup U_2} + \Lambda_{U_1 \cap U_2}$

We claim that  $\mathcal{S}$  is endowed with a unique Borel measure  $\mu$  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  $\Lambda_U(c) \cong \Lambda_U(\rho_c)$ .  $\Lambda = \{\Lambda_U : \mathcal{C}(\mathcal{S}) \rightarrow \mathbb{R}\}_{U \subseteq \mathcal{S}}$  is a family of positive linear functionals such that  $\forall U \subseteq \mathcal{S} \Lambda_U = \Lambda_{\text{int}(U)}$  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 the extension X.2 of the Riesz representation theorem for linear functionals there exists a unique Borel measure  $\mu$  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  $\mathbb{R}$ .  $\forall c \in \mathcal{C} \exists! \rho_c : \mathcal{S} \rightarrow \mathbb{R}$  returning the amount of material for each state  $s \in \mathcal{S}$ . As  $|\Lambda_U(c)| < \infty \forall c \in \mathcal{C} \forall U \subseteq \mathcal{S}$ , then  $\rho_c \in \mathcal{L}(\mathcal{S}, \mu) \equiv \{\rho : \mathcal{S} \rightarrow \mathbb{R} \mid |\int_U \rho d\mu| < \infty \forall U \subseteq \mathcal{S}\}$ .  $\mathcal{L}(\mathcal{S}, \mu) = L^1(\mathcal{S}, \mu)$  by X.3. Let  $\varrho : \mathcal{C} \rightarrow L^1(\mathcal{S}, \mu)$  such that  $\varrho(c) \mapsto \rho_c$ .  $\varrho$  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.  $\square$

<sup>12</sup> Here measure is intended in the mathematical sense (a real valued function of a  $\sigma$ -algebra) which is distinct from other connotations in physics or probability theory.

### 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  $\rho_c$  a function not just of the state  $s$ , 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.

Moreover, the density itself has to be differentiable. The distribution  $\rho_c$  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  $C^1(\mathcal{S}) \equiv \{\rho : \mathcal{S} \rightarrow \mathbb{R} \mid \rho \text{ is differentiable}\}$ .

*Justification.* We claim  $\mathcal{S}$  is a differentiable manifold. Let  $\rho_c : \mathcal{S} \rightarrow \mathbb{R}$  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 in  $U \subseteq \mathcal{S}$ . 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  $\rho_c$  transforms as a density.  $\rho_c(s) = \rho_c(q^i(s)) = \rho_c(\hat{q}^j(s))$ . 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  $\rho_c$  be the distribution associated with state  $c$ . Let  $\rho_c$  be strictly monotonic in a region  $U \subseteq \mathcal{S}$ . Then its fibers in  $U$ , the inverse images of the values, are connected hyper-surfaces. 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  $\delta_i^1$  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  $\rho_c$  be the distribution associated with state  $c$ . Let  $\rho_{c_0}$  be the distribution associated with state  $c_0$ . Let  $\rho_{c_0}$  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  $U$ .  $\rho_{c_0}$  and  $\delta\rho$  are differentiable over  $U$  as they are strictly monotonic over  $U$ .  $\rho_c$  is a linear combination of differentiable functions over any arbitrary region  $U$ .  $\rho_c$  is differentiable.

We claim  $\mathcal{C} \cong C^1(\mathcal{S}) \cap L^1(\mathcal{S}, \mu)$ . Let  $\rho_c$  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  $\exists U \in \mathcal{S} \mid \int_U \rho_c d\mu \neq \int_U \rho d\mu$ . Both integrals are finite as  $\rho_c, \rho \in L^1(\mathcal{S}, \mu)$ . Both integrals are expressible with the same state variable 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  $c_1 \in \mathcal{C} \mid \varrho(c_1) = \rho$ .  $\varrho : \mathcal{C} \rightarrow C(\mathcal{S}) \cap L^1(\mathcal{S}, \mu)$  is surjective.  $\varrho$  is a homomorphism, as justified in V.2.  $\varrho$  is an isomorphism between  $\mathcal{C}$  and  $C(\mathcal{S}) \cap L^1(\mathcal{S}, \mu)$  as vector spaces.  $\square$

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 discontinuous 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.<sup>13</sup>

Another consideration is that while the norm associated to  $L^1(\mathcal{S})$  is  $\int_{\mathcal{S}} |\rho| d\mu$ , the expression  $\int_{\mathcal{S}} \rho d\mu$  is perhaps more physically meaningful. 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  $\langle f, g \rangle = \int_{\mathcal{S}} f g d\mu$ . 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:

$$\begin{aligned}\Lambda_{\mathcal{S}}(c) &= \int_{\mathcal{S}} \rho_c d\mu = \langle 1, \rho_c \rangle \\ \Lambda_U(c) &= \int_U \rho_c d\mu = \langle 1_U, \rho_c \rangle \\ \rho_c(q_0, p_0) &= \int_{\mathcal{S}} \delta_{q_0, p_0} \cdot \rho_c d\mu = \langle \delta_{q_0, p_0}, \rho_c \rangle \\ p_{total}(c) &= \int_{\mathcal{S}} p \cdot \rho_c d\mu = \langle p, \rho_c \rangle \\ q_{avg}(c) &= \frac{\int_{\mathcal{S}} q \cdot \rho_c d\mu}{\int_{\mathcal{S}} \rho_c d\mu} = \frac{\langle q, \rho_c \rangle}{\langle 1, \rho_c \rangle}\end{aligned}$$

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 over  $\mathcal{S}$ . But these vectors  $(1_U, p, \dots)$  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  $\rho_c$  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  $1 \times 1$  matrix, there can only be one  $k$ . This gives  $\partial_k \hat{k} = d_q q$  and  $\hat{k} = d_q q k + f(q)$ . 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  $\hat{k} = d_q q k$  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 possibilities for  $q$  and  $k$  will give us  $\hbar$  possibilities.<sup>14</sup> The cardinality for states 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

<sup>13</sup> One can, though, make the argument that given any distribution  $\rho$  one can find an infinitely smooth  $\rho_{sm}$  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 makes this mathematically well defined.

<sup>14</sup> 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  $\hbar dk = dp = m du$  between kinetic and conjugate momentum, derived later, sets the relationship between units.

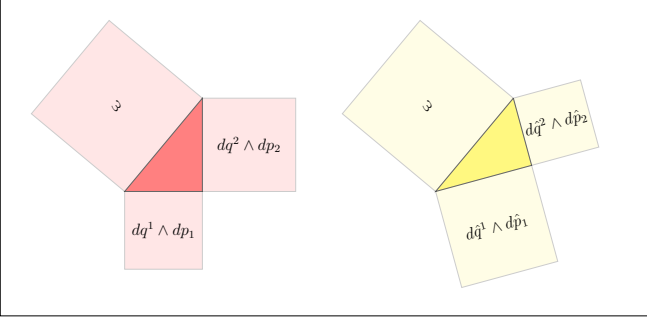


FIG. 2. The metric  $\omega$  and its components along two independent degrees of freedom. As we change state variables, the new d.o.f. are also independent and therefore orthogonal to each other. Note: this is a 2D conceptualization of 4D space.

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 that is preserved as we change state variables. Each pair  $(q^i, k_i)$  is expressed by an independent unit with the cardinality 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 number of states 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 possibilities 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 cardinality of possibilities 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 two-form  $\omega = \sum dq^i \wedge dp_i$  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 cardinality 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  $\theta = k dq$  represents the geometrical object we associate

with each particle state. The canonical two-form  $\omega$  is the metric 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 metric allows us to measure the cardinality of 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  $\omega$ .

*Justification.* We claim the simplest 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  $i = 1 \dots n - 1$  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.  $|\partial_i \hat{k}^j| = d_q q$ .  $\partial_i \hat{k}^j$  is one dimensional as its components must be fully specified by the previous relationship. Let  $k = k^1$ .  $\partial_k \hat{k} = d_q q$ .  $\hat{k}(q, 0) = 0$  as a change in units does not change the 0 value.  $\hat{k} = d_q q k$ .  $k$  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  $\mathcal{Q}$  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  $\omega$  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  $\omega$  is the canonical two form.  $\omega$  is invariant under state variable change.  $\omega$  is a symplectic metric for  $\mathcal{S} = T^*\mathcal{Q}$ .  $\mathcal{S} = (T^*\mathcal{Q}, \omega)$  is a symplectic manifold.

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  $\omega$  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  $\omega$  such that  $\int_{U \subset \mathcal{S}} \rho_c \omega$  where  $U$  is any two dimensional surface in  $\mathcal{S}$ .  $\omega$  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.  $\square$

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  $\mu$  for the cardinality 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 has 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.<sup>15</sup>

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.<sup>16</sup> 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 both a topological space and a measure expressible in terms of state variables that allows us to count states and possibilities. We recovered the metric 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 metric typical of classical phase space. In short: being able to measure the amount of material is what allows us to measure the cardinality of 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.

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

<sup>15</sup> One may still posit that space is fundamentally discrete at scales where both classical and quantum mechanics cease to be valid. For that to work, the measure defined on such a discrete space has to become the measure on the classical phase space. That is: it must be a discrete version of  $T^*\mathcal{Q}$  and not  $\mathcal{Q}$  by itself.

<sup>16</sup> 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.

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
- The system 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  $(\hat{b}) \rightarrow (a, b)$  cannot be an injection, therefore we arrive at a contradiction.

Of those three requirements we can only pick two. If we want determinisms 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.<sup>17</sup> We can always find such 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 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

<sup>17</sup> Also note that any Hamiltonian is locally isomorphic to a free particle (citation needed)

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.<sup>18</sup>

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.  $\square$

## 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  $\mathcal{J}$  has to end up in  $\hat{\mathcal{J}}$ . In math terms  $\rho_{\hat{\mathcal{J}}}(\hat{\mathcal{J}}) = \rho_{\mathcal{J}}(\mathcal{J})$ .

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 best way to look at it is that initial and final sets of states have to possess the same cardinality of states and possibilities, which were defined by our metric  $\omega$ . 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 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 metric  $\omega$ .

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

*Justification.* We claim  $\mathcal{J}_{\Delta t}$  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.  $\mathcal{J}_{\Delta t}$  is a deterministic and reversible evolution map and by IV.3 is a self-homeomorphism.

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

If we assume continuous time evolution we have the following:

**Proposition V.7.** A continuous deterministic and reversible process 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

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

*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 space  $\mathcal{S}$  for the particles of a classical material is a symplectic manifold by V.5. The map for infinitesimal time evolution  $\mathcal{J}_{dt}$  is an infinitesimal self-symplectomorphism by V.6. By X.4 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  $b = \{1, \dots, n\}$  we have  $d_t p_i(-1) = \partial_{q^i} H$ . For  $b = \{n+1, \dots, 2n\}$  we have  $d_t q^i(+1) = \partial_{p_i} H$ .  $\square$

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

<sup>18</sup> 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.



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, 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, 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 set of states 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 process: the Hamiltonian  $H : T^*\mathcal{Q} \rightarrow \mathbb{R}$  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  $d_t \hat{t} d_{\hat{t}} q^i = \partial_{p_i} H$ .  $q^i : T^*\mathcal{Q} \rightarrow \mathbb{R}$  and  $H : T^*\mathcal{Q} \rightarrow \mathbb{R}$  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  $\mu$  needs to be modified to define integrals over the time variable.

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 cases outlined are pretty 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 is an element  $s \in \bar{\mathcal{S}}$  while a time evolution is a line  $\bar{s} \subset \bar{\mathcal{S}}$  that passes once

at all times. Specifying an evolution map means giving a set of lines such that each state is traversed once and only once. Locally, each state will be given by  $(q^i, k_i, t)$  therefore  $\bar{\mathcal{S}}$  is a  $2n + 1$  dimensional manifold. A region at equal time will be a hyper-surface  $\mathcal{S}_{t=t_0} \subset \bar{\mathcal{S}}$  that includes all possible configurations at a particular time. Such hyper-surface will need to cut across all evolutions once and only once, therefore not all functions  $t : \bar{\mathcal{S}} \rightarrow \mathbb{R}$  are suitable time variables.

**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 hyper-surface 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  $s \in U$  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 hyper-surface of  $\bar{\mathcal{S}}$ . Let  $t : \bar{\mathcal{S}} \rightarrow \mathbb{R}$  be a time variable.  $\mathcal{S}_{t=t_0} = \{s \in \bar{\mathcal{S}} \mid t(s) = t_0\}$  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 hyper-surface.  $\square$

For the composite state we need to take a different approach. A distribution of material over states at equal time  $\rho_t : \mathcal{S}_{t=t_0} \rightarrow \mathbb{R}$  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 *evolution space* of a classical material, which we'll indicate by  $\bar{\mathcal{C}}$ . To each evolution  $\bar{c} \in \bar{\mathcal{C}}$  will correspond a distribution  $\rho_{\bar{c}} : \bar{\mathcal{S}} \rightarrow \mathbb{R}$  over all particle states at all times. This distribution will be differentiable: it's differentiable over all equal time hyper-surfaces 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 hyper-surfaces: no matter what time variable we choose, the amount of material found at a particular time has to be finite.

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

*Justification.* We claim  $\bar{\mathcal{C}} \cong G \subseteq C^1(\bar{\mathcal{S}}) \cap L_t^1(\bar{\mathcal{S}}, \mu)$ . Let  $\rho_{\bar{c}} : \bar{\mathcal{S}} \rightarrow \mathbb{R}$  be the distribution of material associated with an evolution  $\bar{c} \in \bar{\mathcal{C}}$ . Let  $t : \bar{\mathcal{S}} \rightarrow \mathbb{R}$  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.  $\rho_{\bar{c}} \in L_t^1(\bar{\mathcal{S}}, \mu)$  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  $\bar{\mathcal{C}} \cong C^1(\bar{\mathcal{S}}) \cap L_t^1(\bar{\mathcal{S}}, \mu)$ . 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.  $\square$

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  $\bar{s} \subset \bar{\mathcal{S}}$ : 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  $\bar{s}$ . 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 hyper-surface at constant time.

### VI.C. Relativistic cotangent bundle

We now want to find the metric for our 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 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.<sup>19</sup>

As in the time independent case, we are interested in invariant densities which are defined on co-vectors.

<sup>19</sup> 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.

The co-vector  $k_i dq^i$  is completed with a time component  $\bar{\omega} dt$ .<sup>20</sup> 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  $\rho_c$  and therefore by the measure  $\mu$  and the metric  $\omega$ .

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 hyper-surface 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 metric 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 the metric is defined and its projections on the planes defined by each d.o.f., similar to the multiple d.o.f.:

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

But in the previous case, the right angle was between the two independent d.o.f.. In this case, the right angle is between the metric 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 metric  $\omega$  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. The metric allows us to

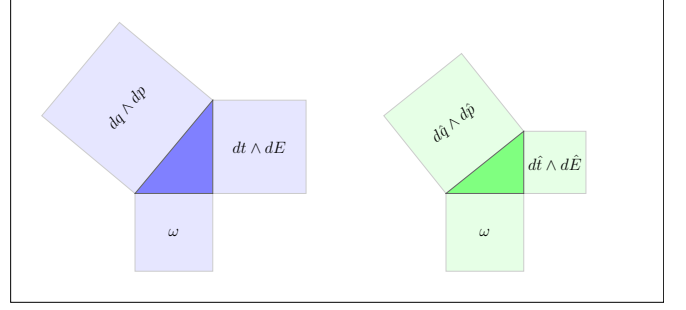


FIG. 3. The metric  $\omega$  and its components along temporal and standard degrees of freedom. The two d.o.f. are not independent and therefore not orthogonal. The temporal d.o.f. is orthogonal to the metric as states are defined (and counted) independently of time. Compare with FIG. 2. Note: this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

count the possibilities within a degree of freedom, adjusting the metric on  $T^*\mathcal{M}$  to avoid the counting problems introduced by time variable changes.

The metric actually fixes two other problems. The first is: how do we 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 second problem is: how do we prevent time variable changes that disturb determinism and reversibility? The invariance of  $\omega$  already guarantees that. As we saw before, the fact that the vector product is conserved within a degree of freedom is equivalent to saying that changing variables does not change the possibilities within a d.o.f. In fact, the infinitesimal displacement  $d_s x^i$  for a general infinitesimal transformation that preserves the vector product on a surface is a divergence free field. The conservation of the Minkowski product across the temporal and standard d.o.f. means that changing variables does not introduce rotation between time and state variables. In fact, the infinitesimal displacement  $d_s x^i$  for a general infinitesimal transformation that preserves the Minkowski product on a surface is a curl free field.

TODO: add proofs in the appendix

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 metric  $\omega$  allows us to measure the cardinality of possibilities.

**Proposition VI.3.** The complete state space  $\bar{\mathcal{S}}$  for the particles of a classical material is a hyper-surface of the

<sup>20</sup>  $\bar{\omega}$  is the classical analogue of the wave frequency. We use  $\bar{\omega}$  to distinguish from the phase-space metric  $\omega$ .

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  $\omega$  be the metric. The components of  $\omega$  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 hyper-surface 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  $f : T^*\mathcal{M} \rightarrow \bar{\mathcal{S}} \mid (q^i, p_i, t, E) \mapsto (q^i, p_i, t)$ .  $f$  is an open map. The topology of  $\bar{\mathcal{S}}$  is the subspace topology. Let  $\rho : T^*\mathcal{M} \rightarrow \mathbb{R}$  be an invariant density.  $\rho$  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.  $\square$

#### 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 the metric as the number of possibilities 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 a homogeneous classical material is a self-symplectomorphism on  $T^*\mathcal{M}$ . That is:  $\mathcal{T}_{\Delta s} : T^*\mathcal{M} \rightarrow T^*\mathcal{M}$  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.  $\omega$  is the metric for the cardinality of possibilities.  $\omega$  is invariant under

a deterministic and reversible evolution map.  $\mathcal{T}_{\Delta s}$  is a symplectomorphism by definition.  $\square$

If we assume continuous time evolution we have the following:

**Proposition VI.5.** A continuous time-dependent deterministic and reversible process for a homogeneous classical material admits an invariant Hamiltonian  $\mathcal{H} : T^*\mathcal{M} \rightarrow \mathbb{R}$  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 X.4 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  $a = \{1, \dots, n\}$  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  $a = \{n+2, \dots, 2n+1\}$  we have  $d_s p_i(-1) = \partial_{q^i} \mathcal{H}$ .  $\square$

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  $\lambda : T^*\mathcal{M} \rightarrow \mathbb{R}$  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.<sup>21</sup>

**Proposition VI.6.** Let  $\mathcal{T}_{\Delta s} : T^*\mathcal{M} \rightarrow T^*\mathcal{M}$  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  $d_s t < 0$ .

<sup>21</sup> This represents the classical analogue for quantum anti-particle states.

*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  $s_1, s_2 \in T^*M$  be two physical states connected by deterministic and reversible evolution.  $\text{sign}(d_s t(s_1)) = \text{sign}(d_s t(s_2))$ . Let  $U \equiv \{s \in T^*M \mid d_s t(s) > 0\}$  and  $V \equiv \{s \in T^*M \mid d_s t(s) < 0\}$ .  $U \cap V = \emptyset$ . Let  $\gamma : [s_0, s_1] \rightarrow T^*M$  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  $\gamma([s_0, s_1]) \subseteq V$ .  $\square$

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:  $\mathcal{H} = \lambda(t, q^i, E, p_i) (H(t, q^i, p_i) - E - h(t))$  where  $\lambda = d_s t$ ,  $h(t)$  is an arbitrary function and, most importantly,  $E = H(t, q^i, p_i) + h(t)$ .

The arbitrary function  $h(t)$  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  $s \in \bar{S} \subset T^*M$ , 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  $T^*M$  such that  $\mathcal{H} \neq 0$  outside of the complete state space  $\bar{S}$ . This way we can directly use the constraint  $\mathcal{H} = 0$  to identify  $\bar{S}$  as a subspace of  $T^*M$ .

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  $\mathcal{H} = \lambda(t, q^i, E, p_i) (H(t, q^i, p_i) - E)$  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  $\lambda : \bar{S} \rightarrow \mathbb{R}$  where  $\lambda(\xi^a) = d_s t$ .  $\bar{S}$  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  $\lambda : T^*M \rightarrow \mathbb{R}$  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  $\mathcal{H} = \lambda(t, q^i, E, p_i) (H(t, q^i, p_i) - E + h(t, q^i, E, p_i))$  without loss of generality. We have  $\partial_{q^i} \lambda (H - E + h) + \lambda (\partial_{q^i} H + \partial_{q^i} h) = \lambda \partial_{q^i} H$ . This holds for all choices of  $t$ . In particular for  $t = s$ ,  $\lambda (\partial_{q^i} H + \partial_{q^i} h) = \lambda \partial_{q^i} H$ .  $\partial_{q^i} h = 0$ .  $h$  is not a function of  $q^i$ . Repeat the same procedure for  $p_i$ .  $h$  is not a function of  $p_i$ .  $\lambda = d_s t = -\partial_E \mathcal{H} = -\partial_E \lambda (H - E + h) - \lambda (-1 + \partial_E h)$ . In particular for  $t = s$ ,  $\lambda = -\lambda (-1 + \partial_E h)$ .  $\partial_E h = 0$ .  $h$  is not a function of  $E$ .  $h$  is only a function of  $t$ . We can redefine  $H = H + h$  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  $\lambda$ .  $H - E = 0$ .

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

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

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

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

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  $\gamma : \mathbb{R} \rightarrow T^*M$  be a parametrization for a particle evolution  $\bar{s}$ . 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  $\bar{C}$  is not simply the Lebesgue integrable differentiable functions, but those that are integrable on particular hyper-surfaces. The complete state space  $\bar{S}$  is a hyper-surface 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  $\bar{\gamma}$  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 re-impose 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 mechan-

ics. 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 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.  $\square$

### 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  $x^\alpha$  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 hyper-surface 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,  $u^\alpha = d_s x^\alpha = d_{\hat{s}} \hat{s} d_{\hat{s}} x^\alpha = d_{\hat{s}} \hat{s} \hat{u}^\alpha$ , 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 hyper-surface 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 parti-

cle 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 hyper-surface 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 hyper-surface of the tangent bundle  $T\mathcal{M}$ .

*Justification.* We claim the complete state space is a hyper-surface of the cotangent bundle of the space-time manifold,  $\bar{\mathcal{S}} \subset T^*\mathcal{M}$ . Let  $\bar{\jmath} \in \bar{\mathcal{S}}$  be the evolution of a particle for a minimal classical material. Let  $f(\bar{\jmath}) \mapsto x^\alpha(s)$  be the function that returns the space-time trajectory associated with the evolution.  $f$  is injective as the material is minimal.  $\bar{\jmath}$  can be identified by a trajectory  $x^\alpha(s)$  which can be fully described by coordinates defined on  $\mathcal{M}$ . Let  $\jmath \in \bar{\mathcal{S}}$  be the particle state for a minimal classical material. Let  $f(\bar{\jmath}, t) \mapsto \jmath$  the function that returns the state at a particular time  $t$  along an evolution  $\bar{\jmath}$ . Such function is a bijection as the evolution is deterministic and reversible.  $\jmath$  can be identified by an evolution  $\bar{\jmath}$  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 hyper-surface of  $T\mathcal{M}$ . Let  $\jmath \in \bar{\mathcal{S}}$  be the particle state for a minimal classical material. Let  $x^\alpha(s)$  be the evolution associated with that state. Let  $t(\jmath)$  be the instant of time at which  $\jmath$  is defined.  $q^\alpha(\jmath) \equiv x^\alpha(s(t(\jmath)))$  are state variables as they are real function of the state.  $q^\alpha$  is differentiable therefore  $d_s x^\alpha$  exists. Suppose  $d_s x^\alpha$  is not part of the state.  $d_s x^\alpha = d_s q^\alpha$  is determined by the law of evolution.  $x^\alpha(s)$  is determined by the law of evolution.  $q^\alpha$  determines the trajectory  $x^\alpha(s)$  and the state  $\jmath$ , as the material is minimal. This is a contradiction:  $\jmath$  is not identified by just  $q^\alpha$ .  $u^\alpha = d_s x^\alpha$  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 homomorphism between  $T^*\mathcal{M}$  and  $T\mathcal{M}$ .  $\bar{\mathcal{S}}$  is a hyper-surface of  $T^*\mathcal{M}$  therefore its image is a hyper-surface of  $T\mathcal{M}$ .

We claim  $(x^\alpha, u^\alpha) \mapsto (q^\alpha, p_\alpha)$  is a diffeomorphism. Let  $\bar{\tau}$  be the evolution of a minimal classical material. Let  $\rho_{\bar{\tau}} : \bar{\mathcal{S}} \rightarrow \mathbb{R}$  be the distribution associated with  $\bar{\tau}$ .  $\rho_{\bar{\tau}}(\jmath(x^\alpha, u^\alpha))$  as the material is minimal.  $\rho_{\bar{\tau}}(x^\alpha, u^\alpha)$  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.  $\square$

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  $p_\alpha$ . Since  $\partial_{p_\alpha} \mathcal{H} = d_s q^\alpha = u^\alpha$ ,  $\mathcal{H}$  is convex in  $p_\alpha$ . 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  $v^i = dt x^i$ .

**Proposition VII.2.** Let  $\mathcal{H}$  be an extended Hamiltonian defined on  $\mathbf{T}^*M$ . 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  $p_\alpha$ . 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  $q^\alpha$ ,  $u^\alpha$  must be invertible. The components  $u^\alpha$  are monotonic in  $p_\alpha$ .  $\partial_{p_\alpha} \mathcal{H} = d_s q^\alpha = u^\alpha$ . The derivatives of  $\mathcal{H}$  in  $p_\alpha$  are monotonic in  $p_\alpha$ .  $\mathcal{H}$  is convex in  $p_\alpha$ .

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

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. There must exist a metric  $g$ , consistent with  $\omega$ , that allows one to calculate the cardinality of possibilities for initial conditions. That is:  $\omega(dq^\alpha, dp_\alpha) \propto g(dx^\alpha, du^\alpha)$  where the arguments are the value ranges for the respective quantities.

Since both  $dx^\alpha$  and  $du^\alpha$  are four vector components and the metric must be linear,  $g$  is a rank two tensor on  $M$ . To each direction in space-time  $dx^\alpha$  is associated a direction in the tangent space of velocities  $du^\alpha$  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  $dx^\alpha$  and  $du^\beta$  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  $x^\alpha$  is a choice of units for  $u^\alpha$  and therefore they cannot belong to different d.o.f. This means  $g$  is symmetric.  $M$  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 possibilities over the temporal d.o.f. into the possibilities 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  $M$  is a pseudo-Riemannian manifold  $(M, g)$ .  $g$  has signature  $(3, 1)$  and locally is  $g = (dx^i)^2 - c^2 dt^2$ .

*Justification.* N.B. Though  $du^\alpha \in TTM$ , its components change as the components of vectors in  $TM$  under

coordinate transformation. There is a canonical way to associate between elements of  $TTM$  and  $TM$ . Therefore we consider  $du^\alpha \in TM$  a vector as it is standard in physics.

We claim  $(M, g)$  is a pseudo-Riemannian manifold where  $g$  is a metric induced by  $\omega$  under kinematic equivalence. Let  $(dx^\alpha, du^\alpha)$  be an infinitesimal range of possibilities for the kinematic variables of the particles of a minimal classical material. There exists a map  $g : TM \times TM \rightarrow \mathbb{R}$  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.  $(M, g)$  is a pseudo-Riemannian manifold.

We claim  $g$  is locally Minkowskian. Let  $x^\alpha$  be a choice of coordinate that diagonalizes  $g$ .  $x^\alpha$  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  $(3, 1)$ . 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.  $\square$

The metric tensor  $g$  and the constant  $c$  found here are indeed the ones used in the context of relativity. Yet, they have acquired a different meaning.  $g$  does not measure space-time distances: it measure cardinality of possibilities, like  $\omega$ . Directions in space are orthogonal not because the angle between them is  $90^\circ$ , 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  $\omega$ . 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 cardinality 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 conceptually 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  $\omega$  and therefore  $g$ . Non-local quantum effects, such as the ones featured in EPR experiments, are not examples of deterministic and reversible motion. Therefore they do not have anything to violate: the maximum speed does not apply to them.

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

The link between  $\omega$  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  $dp_\alpha = mdu_\alpha$ , where  $m$  is the constant of proportionality. We recognize  $m$  as the inertial mass. Integrating the expression, we have  $p_\alpha = mu_\alpha + A_\alpha(x)$  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) for which the modulus of the velocity is constant. We have  $u^\alpha u_\alpha = -c^2 = \frac{1}{m^2}(p^\alpha - A^\alpha)(p_\alpha - A_\alpha)$ .  $\mathcal{H} = \frac{1}{2m}((p^\alpha - A^\alpha)(p_\alpha - A_\alpha) + m^2 c^2)$ . 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  $\mathcal{H} = \frac{1}{2m}((p_\alpha - A_\alpha)g^{\alpha\beta}(p_\beta - A_\beta) + m^2 c^2)$ .

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

We claim the Hamiltonian is  $\mathcal{H} = \frac{1}{2m}((p_\alpha - A_\alpha)g^{\alpha\beta}(p_\beta - A_\beta) + m^2 c^2)$ .  $\frac{1}{m}g^{\alpha\beta}(p_\beta - A_\beta) = u^\alpha = d_s x^\alpha = d_s q^\alpha = \partial_{p_\alpha} \mathcal{H}$ . 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.  $\mathcal{H} = \frac{1}{2m}((p_\alpha - A_\alpha)g^{\alpha\beta}(p_\beta - A_\beta) + m^2 c^2)$ . □

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 equation valid only for massive particles.

We have seen that the inertial mass  $m$ , the invariant speed  $c$  and the reduced Plank constant  $\hbar$  all serve a similar 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 to keep proper book keeping when mapping sets of states. Using natural units means measuring all these quantities through the cardinality of possibilities. 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  $x^\alpha$  already determines the others ( $x^\alpha \rightarrow u^\alpha$  and  $x^\alpha \rightarrow q^\alpha \rightarrow k_\alpha \rightarrow p_\alpha$ ) therefore the link between units of  $u^\alpha$  and  $p_\alpha$  is not a new unit conversion.

Another important thing to notice 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  $x^\alpha$  and  $mu_\alpha$  as state variable. 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  $x^\alpha$  and  $mu_\alpha$  does not change the physics. Expressing the Hamiltonian and the Poisson brackets in terms of the kinematic variables we have:

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

Note how the 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 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  $\tau$  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:

$$\pm d_\tau(\pm md_\tau x^\alpha) = md_\tau^2 x^\alpha = \pm F_{\beta}^\alpha d_\tau x^\beta$$

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-state as they follow the trajectories in the opposite direction. But if we apply parity  $P$  and time reversal  $T$ , changing  $x^\alpha$  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.

To sum up, the kinematic assumption has more inherent consequences 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 trajectories on  $T\mathcal{M}$  with the space of states on  $T^*\mathcal{M}$ . Such link allows state space metric  $\omega$  to induce the space-time metric  $g$ . It allows us to define Lagrangians. It constrains the motion to massive particles under scalar and vector potential forces. Moreover, it allowed to re-understand 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 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

In principle, we could simply introduce assumption ?? and work our way to quantum mechanics. Unfortunately, given the history and baggage associated with quantum systems, there is a high chance of being misconstrued. The intuitive pictures people have constructed or inherited are varied, may start from different premises and therefore it is easy to misinterpret what one is saying. Therefore we will first ground in classical mechanics the intuition that will be needed later, making the step into full quantum mechanics a gentler, and hopefully less confusing, one.

We will see that, as a first approximation, quantum particles are more keenly described by classical composite states (i.e. distributions) as, unlike classical particles, they are made of a finite amount of material. We will also see that classical distributions already possess qualitative properties that are typically associated with quantum systems and construct a classical analogue of the uncertainty principle.

### VIII.A. Particles and distributions

As we saw in the previous sections, classical particles should not be understood as point-like objects of finite mass as this would open a series of problems and inconsistencies. The issues created by such misidentification are only amplified within the context of quantum mechanics. In fact, many issues preventing a satisfying intuitive picture stem from this confusion.

The problem goes along these lines. One assumes, implicitly or explicitly, that quantum particles are point-like. But they don't always behave like one, they also behave like a wave, a distribution. During measurements, this distribution coincides experimentally with a probability distribution. So the state of a quantum particle is a probability distribution. Except that it self-interacts. So it's a special quantum probability distribution as the quantum particle behaves like it's in more than one place at once. We learn this picture and, sometimes begrudgingly, we get used to it.

Here we contend that this picture is not only confusing and unnecessary, it also prevents us from understanding fully the deep links between classical and quantum mechanics, in the same way that thinking of classical particles as point-like objects prevented us from seeing a more holistic picture in classical mechanics.

The first thing to note is that all physical objects behave like they are in more than one place at once. The earth is both under Milano and New York. Your right index finger is not, hopefully, at the same position of your left eye. In fact, based on what we saw before, any object of finite mass is a distribution over a region, as discontinuous distributions are ruled out. 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 behaves in a different way than all other physical objects. Shouldn't *he* have the burden of proof?

Part of the confusion seems to be the idea that, since the electron is an elementary particle, it should have no spatial extent. But that does not follow. Lack of spatial extent and lack of substructure are independent concepts. In fact: we can construct an infinitesimal object that is not elementary. Suppose we have a sphere with dots moving on its surface. All their dynamics can be expressed by their polar and azimuthal angles. If we scale the sphere, the dynamics stays the same. We can shrink the whole sphere to a point: it is now a point-like object that is not elementary, as there is motion on its surface.

Not only point-like objects don't have to be elementary, but it's not clear 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, which is an additional structure. If everything is just points, where does this come from?

Entertaining the thought that an electron is actually physically distributed in space may seem preposterous, probably because it's not how we are taught to think about it. But it's no more preposterous than to think it point-like, described by a probability distribution that self-interacts.

The first payoff of accepting that a quantum particle is a finite amount of material distributed in space is that this is more consistent with classical mechanics as we have defined it<sup>22</sup>. The second payoff is that some features of classical mechanics already explain part of the behavior that is usually thought as inherent to quantum mechanics.

### VIII.B. Classical uncertainty principle

Suppose we have a finite amount of classical material distributed over a single d.o.f.  $a_m = \int \rho_c dq \wedge dp$ . The Shannon entropy associated with the distribution, which can be thought as the amount of information required to identify a particle of the distribution, is finite as well.  $I(\rho_c) = \int \rho_c \log(\rho_c) dq \wedge dp = I_0$ . TODO: prove in appendix.

As the state evolves into its final state, the amount of material remains the same:  $a_m = \int \rho_c dq \wedge dp$ . The final entropy will also remain the same since the entropy, like the distribution itself, is just another integrable function:  $I(\rho_f) = I_0$ .

In the same way that each particle is constrained to move along states at equal energy, the composite state is constrained to move along distributions at equal entropy. This makes sense: if the motion is deterministic and reversible, the same amount of information is required to identify a particle of material.

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  $a_m$  and the total entropy  $I_0$  and find the distribution that minimizes the product of the variance

---

<sup>22</sup> It is also more consistent with general relativity.

$$\sigma_q^2 \sigma_p^2 = \int (q - \mu_q)^2 \rho dq dp \int (p - \mu_p)^2 \rho dq dp.$$

$$\begin{aligned} L &= \int (q - \mu_q)^2 \rho dq dp \int (p - \mu_p)^2 \rho dq dp \\ &\quad + \lambda_1 \left( \int \rho dq dp - \rho_{tot} \right) \\ &\quad + \lambda_2 \left( - \int \rho \ln \rho dq dp - H_0 \right) \\ \delta L &= \int \delta \rho [(q - \mu_q)^2 \sigma_p^2 + \sigma_q^2 (p - \mu_p)^2 \\ &\quad + \lambda_1 + \lambda_2 \ln \rho + \lambda_2] dq dp = 0 \\ \rho &= \frac{\rho_{tot}}{\sqrt{4\pi^2 \sigma_q^2 \sigma_p^2}} e^{-\frac{(q - \mu_q)^2 (p - \mu_p)^2}{4\sigma_q^2 \sigma_p^2}} \\ &= \frac{\rho_{tot}}{e^{H_0 - 1}} e^{-\frac{\pi^2 (q - \mu_q)^2 (p - \mu_p)^2}{\exp(2(I_0 - 1))}} \end{aligned}$$

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 state. This can be addressed with some extra assumptions.

Suppose the evolution of the distribution is deterministic and reversible but that each particle is undergoing non deterministic motion. Suppose  $\rho$  is a distribution at equilibrium of such motion. While each particle moves randomly, the overall distribution  $\rho$  stays the same: the non deterministic motion is effectively just reshuffling the particles within the same distribution. The value of the entropy of the distribution is set by the non deterministic motion of the particles: the stronger it is, the greater the entropy.

Suppose now  $\rho$  evolves adiabatically. At each instant we can assume  $\rho$  to be at equilibrium. At each instant the particles are reshuffled but the amount of entropy of the distribution remains the same. The overall evolution is still Hamiltonian, but the trajectories predicted by the framework are now the average motion of the particles.

If we assumed that all materials were interacting with the same background that caused all particles to exhibit a random walk of the same magnitude, as defined by the information entropy, then the spread of all distributions for all material will satisfy the same inequality. This gives us the classical uncertainty principle:  $\sigma_q \sigma_p \geq \exp(I_\emptyset - 1)/2\pi$  where  $I_\emptyset$  is the entropy associated with the background.

Assuming such a background is not far fetched. In fact, it serves as a remainder of the issues mentioned in assumption IV: as determinism and reversibility at all scales conceptually does not work, there exists a minimum constant interaction to all particles: the system is not perfectly isolated. If the entropy of our distribution is comparable to the entropy introduced by the background, we are likely in the range where classical mechanics is not valid as the random motion of each particle is not negligible.

The generalization to multiple d.o.f. is outside of the scope of this work. Note, though, that this result is similar in spirit 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.

It should now be clear how in classical mechanics we can already find 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 are general features of an evolution that is deterministic, reversible and adiabatic (i.e. Hamiltonian dynamics). Quantum mechanics, as we'll see, just inherits these traits.

## IX. IRREDUCIBLE SYSTEMS

In this section we turn to a different case in which the relationship between the whole material and its parts is different from the one described in section V. The system is still infinitesimally decomposable, made of infinitesimal parts as before, but this time it is irreducible. That is: the state and the evolution of the whole does not tell us anything about the state and the evolution of the parts. As the internal dynamics is undetermined, we'll find that each irreducible element 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 particle*. 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. Most of the time these two concepts overlap, but they are separate as we can see in the following examples.

Consider the 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 into them. Consider a proton. Its state is given by the state of its three valence quarks. Yet, we cannot divide the proton into the three quarks. The proton is reducible to quarks, but 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. While we are agnostic on whether a quantum particle can be divided, we are implicitly assuming that it can be decomposed. Irreducibility describes the relationship between the state of the whole and the state of the parts, therefore parts are implicit in the definition. While these parts must exist, we cannot isolate them and act on them independently: if we were able to do that we could assign them a state.

While we must be able to talk about parts of a quantum particles, 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 is decomposed some part 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.

The proper way to think of a quantized material is therefore the following: a continuum of infinitesimal parts 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 in the quantum particle state. Consider a part of a quantum particle in an initial state. As we cannot track its motion, it can be any part of the particle in its final state. That is, the motion of a part 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 state 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 affect also the part where no force is exerted. We cannot simply interact with the part at a specific region, as this would allow to assign a state to that part.

Wave particle duality should be clear at this point. 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 one unit and cannot be separated.

The most striking feature, though, is that we can predict the functional form for particle decay. Suppose a quantum particle spontaneously decays into two or more particles. Such process cannot be deterministic as the d.o.f. increase, therefore is not determined by the state of the quantum particle. As the decay is spontaneous, it is not triggered by the environment (i.e. external state). It is therefore triggered by the internal unstated dynamic of the parts. As we assumed the internal dynamic 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 dynamic is more stable for some states. The probability of decay is constant in time. 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. Quantum particle and complex vector spaces

Turning these concept into a precise mathematical formulation is tricky as we need to formally characterize the unstated part which is not physically distinguishable. While many previous results do not apply, the classical derivation still can serve as a template.

As in the classical case, the quantum particle is thought to be decomposable: different parts will be in different places, the evolution of the whole is the composition of the evolution of parts. Therefore V.1 applies and the state space  $\Psi$  of an irreducible material is an abelian group.

Note, though, that we can't think of the parts as independent components. As the unstated part is statistical in nature, it will combine as statistical variables do. Suppose  $X$  and  $Y$  are two random variables,  $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y\rho_{X,Y}$ , where  $\sigma^2$  is the variance,  $\sigma$  is the standard deviation, and  $\rho_{X,Y} = \text{COV}(X,Y)/\sigma_X\sigma_Y$  is the Pearson correlation coefficient. As  $-1 \leq \rho_{X,Y} \leq +1$ , the total variance can be greater or smaller than the sum of the independent parts, and the strength of the uncertainty within the unstated part is something physically

distinguishable.

Therefore, while the motion of the unstated part is not itself physically relevant, the correlation between the motion of the unstated part of two combined parts is very relevant. We'll see how interference can be understood in this context: it is constructive if the randomness is correlated while it is destructive if it is anti-correlated.

So, how can we characterize these relationships between different configurations of the unstated parts? We can look at the classical derivation for guidance. In that case, we looked at the set of transformations that increase/decrease the amount of material. Those are the only type of transformation that we can apply on a classical system that do not change the overall state given by the distribution.

As the classical material is homogeneous, permutations between classical particles in the same state  $s$  do not matter: the ordering is unstated. In a way, increasing/decreasing the amount of material increases/decreases the unstated part of the system.

For the irreducible material we follow the same strategy: we find the state transformations that affect the unstated part while keeping the overall state, the distribution, unchanged.

Suppose we have  $\psi_1$  a state for an irreducible material characterized by one d.o.f. The material will be spread over the configurations of the infinitesimal fragments. As the material is homogeneous, the material is spread evenly among the such possible configuration. Counting configurations is therefore equivalent to counting amount of material. Each fragment would be identified by a pair of independent random variables  $(A, B)$ , as a d.o.f. is two-dimensional. The shape of the distribution and the actual values are not physically distinguishable: these are just dummy variables we assign to count the fragment configurations. We choose A and B to be uniform distributions from 0 to 1, so that the area is a square of unit size, which represent the unit of unstated configurations the material is spread over.

Suppose we have another state  $\psi_2$ , where only unstated part is changed. The unstated part can also be characterized by a square formed by two independent uniform distributions (C, D). Counting states on (C,D) is equivalent to counting them on (A,B), as they are simply different arrangements. As all distributions are uniform, they can only be connected by a linear operation. That is:  $C = aA + bB$  and  $D = -bA + aB$ . We can write  $\psi_2 = \tau(a, b)\psi_1$  as  $a$  and  $b$  are the only parameters of the transformation.

The transformation  $\tau(a, b)$  increases/decreases the area by the factor  $a^2 + b^2$  and rotates the square by  $\arctan(b/a)$ . If we set  $c = a + ib$  those same quantities can be expressed as  $\rho^2 = |c|^2$  and  $\theta = \arg(c)$ , the modulus and the phase of a complex number. In other words: complex numbers provide a representation for the group of rotations and magnitude rescaling.

For multiple independent d.o.f., we have one square each. As the material is irreducible, all areas need to

be always the same since relative changes would allow to learn something about the motion of the infinitesimal fragments. This means that the transformation  $\tau(c)$  has to be the same for all d.o.f. In other words, the group of transformations is the same regardless of the number of independent d.o.f.

Mathematically, the state space  $\Psi$  for an irreducible material is a complex vector space.

**Proposition IX.1.** The state space  $\Psi$  for an irreducible material is a vector space over  $\mathbb{C}$ .

*Justification.* TODO □

If  $\mathcal{Q}$  is the space that defines our units, a state  $\psi$  must tell us for each point  $q \in \mathcal{Q}$  how much material is there and the correlation of the unstated part. That is: the state will be identified by  $\psi : \mathcal{Q} \rightarrow \mathbb{C}$ , a complex function over the  $\mathcal{Q}$ .

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 of material. Therefore the function  $\psi \in L^2(\mathcal{Q}, \mu)$  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 particles of the material 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  $\Psi$  such that  $|\psi|^2 = 1$ . Note that changing the phase by the same amount everywhere is only changes the unstated part. Such change does not change the state, therefore distributions whose differ by a total phase are not physically distinguishable and represent the same physical state.

**Proposition IX.2.** The state space  $\Psi$  for an irreducible material is isomorphic to a subspace of the space of Lebesgue square integrable differentiable. That is  $\Psi \cong \Psi(\mathcal{Q}, dp^i) \subseteq C^1(\mathcal{Q}) \cap L^2(\mathcal{Q}, \mu)$  under the isomorphism  $\Upsilon_q : \Psi \leftrightarrow \Psi(\mathcal{Q}, dp^i)$  where  $\Psi(\mathcal{Q}, dp^i)$  is the space of wave functions over  $q^i$ . The state space for a quantum particle is the projective space  $P(\Psi)$ .

*Justification.* TODO □

### IX.C. Inner product and operators

Both in the classical and quantum state the full state is identified by distributions on  $\mathcal{Q}$ . The  $q^i$  do not label the states, but 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 theory of differentiable manifolds and metric spaces. For an irreducible material, the infinitesimal parts cannot be given a state, cannot be identified by a set of local variables. We have to develop other mathematical tools.

To be on par with classical mechanics we need at least address three problems. We need to be able to compare

two quantum 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 state variable distributions and expectations: this will give us linear operators. We need to identify the conjugate quantities  $k_i$  and  $p_i$ . Let's start with the first.

The idea is that, given two states  $\psi_1$  and  $\psi_2$ , we should be able to tell how much of  $\psi_1$  is prepared in the same way of  $\psi_2$ . That is  $\psi_1 = \alpha\psi_2 + \psi_3$  is the combination of the part that is prepared like  $\psi_2$  (e.g. same distributions in position and momentum) and a part that is prepared completely differently. Let  $P_{\psi_2} : \Psi \rightarrow \Psi$  be the operator that gives us the equally prepared part  $P_{\psi_2}(\psi_1) \equiv \alpha\psi_2$ . This operator must be linear in  $\psi_1$ : combining parts that are equally prepared still gives us parts that are equally prepared. It also gives us the same result if applied once or twice: once we isolated the part that is equally prepared, it remains the part that is equally prepared.  $P_{\psi_2}$  is therefore a projection: a linear operator such that  $P_{\psi_2}^2 = P_{\psi_2}$ .

This projection has other useful properties. If  $\psi_1$  is already prepared like  $\psi_2$ , then its projection on  $\psi_2$  will be equal to  $\psi_1$ . Conversely, if the projection of  $\psi_1$  is equal to itself, then it means that the state is already prepared like  $\psi_2$ . That is:  $P_{\psi_2}(\psi_1) = \psi_1 \Leftrightarrow \psi_1 = \alpha\psi_2$ . Moreover, because the material is irreducible, if 23% of  $\psi_1$  is prepared like  $\psi_2$ , then 23% of  $\psi_2$  is prepared like  $\psi_1$ . If not, we could learn more (or less) about  $\psi_1$  by studying the part of  $\psi_2$  that is prepared like  $\psi_1$ . Therefore we have  $\frac{|P_{\psi_2}(\psi_1)|^2}{|\psi_1|^2} = \frac{|P_{\psi_1}(\psi_2)|^2}{|\psi_2|^2}$ . Lastly, if we take the projection of  $\psi_1$  along  $\psi_2$  and we project it again on  $\psi_1$ , we end up with a part of  $\psi_1$ . As this is a proper part of  $\psi_1$ , its state is just  $\psi_1$  rescaled, no change in correlation. That is:  $P_{\psi_1}(P_{\psi_2}(\psi_1)) = a\psi_1$  with  $a \in \mathbb{R}$ .

From the norm and the projection we can define a complex product  $\langle \cdot | \cdot \rangle : \Psi \times \Psi \rightarrow \mathbb{C}$  such that  $P_{\psi_2}(\psi_1) \equiv \frac{\langle \psi_2 | \psi_1 \rangle}{|\psi_2|^2} \psi_2$ . Because of the properties of the norm and the projections, the product is positive definite, linear and conjugate symmetric. This means the state space  $\Psi$  for an irreducible material is an inner product space.

**Proposition IX.3.** The state space  $\Psi$  for an irreducible material is an inner product  $\langle \cdot | \cdot \rangle : \Psi \times \Psi \rightarrow \mathbb{C}$  such that  $\langle \psi_1 | \psi_2 \rangle = \int_{\mathcal{Q}} \psi_1^\dagger(q^i) \psi_2(q^i) dq^i$ .

*Justification.* TODO  $\square$

Note that the inner product only has an indirect physical definition, based on the norm and the projections. We'll therefore want to base our physical intuition on the latter concepts.

Also note that the observations made after V.4 for the classical case are still valid.  $\Psi$  cannot be a complete metric space 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 does no longer represent a physically meaningful entity.

As such, 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 cannot write 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 if  $q$  is a continuous variable. For the same reason, we cannot show that to each state variable is associated an hermitian operator by constructing it from its eigenstates.

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

While  $\mathcal{Q}$  is enough to define the system of units, it cannot be enough to define all the variables over which we can define the distributions. 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 set of state variables that provide another way to subdivide the fragments. Then those variables will form a manifold  $\hat{\mathcal{Q}}$  such that each  $\psi$  can be expressed as a wave function  $\psi(\hat{q})$  over  $\hat{\mathcal{Q}}$ .

As each state  $\psi$  can be identified by a wavefunction on either space, there must be an operator  $\Upsilon_q^q(\psi(q)) = \psi(\hat{q})$  that converts the wave function on one space to the other. As to a linear combination of states corresponds a linear combination of wave function,  $\Upsilon_q^q$  is a linear operator. As converting the space again must give us the original wave function,  $\Upsilon_q^q \Upsilon_q^q = 1$  is unitary. We can now define a have linear operator  $\hat{F}$  corresponding to state variable  $\hat{f} : \hat{\mathcal{Q}} \rightarrow \mathbb{R}$ . Using  $\Upsilon_q^q$  we can express  $\hat{F}$  as an operator over complex functions of  $\mathcal{Q}$  It will be still be linear and hermitian.

TODO: ranges of  $q^i$  and  $\hat{q}^i$  must be the same: there must be a bijective function between them. Bijection is continuous only if  $\hat{q}^i(q^i)$ . Otherwise can be discontinuous.

That is: for each state variable we have a corresponding linear hermitian operator which can be expressed as function multiplication in a suitable choice of  $\hat{\mathcal{Q}}$ .

**Proposition IX.4.** For each fragment variable  $f$  there exists a corresponding self-adjoint linear operator  $F : \Psi \rightarrow \Psi$ . There exist a manifold  $\hat{\mathcal{Q}}$  such that  $f : \hat{\mathcal{Q}} \rightarrow \mathbb{R}$  and  $\Upsilon_{\hat{q}^i} : \Psi \leftrightarrow \Psi(\hat{\mathcal{Q}}, d\hat{q}^i)$  is an isomorphism.

*Justification.* TODO  $\square$

The last thing to do is identify the operators corresponding to  $k_i$  and  $p_i$ , expressed over the space of func-

tions of  $\mathcal{Q}$ . We know that the operators have to be covariant: under a change of variables  $K_q = \partial_q q K_q$ . This is essentially the chain rule: the operator we are looking for is derivation by  $q$ . As it has to be self-adjoint, it will be of the form  $K_q = a \partial_q$  where  $a$  is a real number. We choose  $a$  to be negative by convention and to be unitary so that  $K_q$  is expressed in units of inverse  $q$  as in the classical case. Therefore we have  $K_i = -i \partial_i$  and  $P_i = -i \hbar \partial_i$ .

TODO: Schar

**Proposition IX.5.** The operator  $P_i$  associated with the conjugate quantity of  $q^i$  is  $P_i = -i \hbar \partial_i$ . The operator that changes the variable of the wave function from  $q^i$  to  $p^i$  is  $\Upsilon_{p^i}^{q^i}(\psi(q^i)) = \frac{1}{(\sqrt{2\pi})^3} \int_{\mathcal{Q}} e^{\frac{q_i p^i}{i \hbar}} \psi(q^i) d^n q$ .

*Justification.* TODO □

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

We reached the goal despite  $\Psi$  is not being a complete metric space.

#### IX.D. Irreducibility

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

**Assumption IV** (Infinitesimal reducibility). *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 section ???: 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 tell anything about 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 also derive two types of state evolution: the deterministic and reversible one, corresponding to unitary evolution (e.g. the Schrodinger equation), and a non deterministic one useful to describe measurement interactions, corresponding to the projection postulate. Therefore, as the infinitesimal reducibility assumptions leads to classical mechanics, the irreducibility assumption leads to quantum mechanics.

As we previously saw in section ???, the infinitesimal reducibility assumptions had its problem. How about the irreducibility assumption? How do they compare?

The methodological 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 constituents 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 interacts 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.

The issues rising from a lack of perfect isolation all but disappear. 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 perfect isolation. It is built into the description.

The problem of physically meaningless interaction 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 way, 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 and deterministic evolution at all levels still remains. This will lead us to require a different non deterministic evolution, that will allow us to recover the statistical interpretation of the wavefunction.

The inability to define time by following a deterministic and reversible problem is still present.

There is one additional problem, though, not present in the classical case. The quantum states, as we have defined them, are states at equilibrium: they all provide the same level of description for the fragments; they are all reachable from one another through unitary evolution (which we'll see corresponds to deterministic and reversible evolution). Therefore they cannot and should not be expected to provide a description under non deterministic evolution. 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 a quantum states. We have no reason to expect, though, that the 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 Hilbert space. After all, the state of a single spin one half particle cannot determine the future state of three spin on half particles.

Using unitary evolution in those cases to make predictions, though, may still make physical sense. Quantities that are not affected by the particular evolution of the unstated part will have the same value under non-deterministic evolution and an equivalent quasi-static evolution. The distribution of final states is one such



quantities by the irreducibility assumption: they don't care about the unstated part evolution. But it stands to reason that, at some point, the unstated part will have actual physical consequences. For example: it would be extremely reasonable to expect that is indeed the evolution of the unstated part that decides the moment in which the muon would decay.

In light of what we discussed, the irreducibility assumption is less flawed but it still cannot be fundamental, in the sense that we cannot take it to strictly apply to all of the universe. We should not expect to solve open problems, such as the arrow of time and the measurement problem, within the assumption.  $\square$

### IX.E. Quantum mechanics

We are now ready to write the equation of motion for an irreducible material.

**Proposition IX.6.** A deterministic and reversible evolution map  $\mathcal{T}_{\Delta t} : \Psi \rightarrow \Psi$  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  $\psi_1, \psi_2 \in \Psi$ . The time evolution for the wave function  $\Upsilon_q \mathcal{T}_{\Delta t} \Upsilon^q : S(\mathcal{Q}, dq) \rightarrow S(\mathcal{Q}, dq)$  is a  $C^\infty$ -diffeomorphism.

*Justification.*  $\square$

**Proposition IX.7.** A continuous deterministic and reversible evolution for an irreducible material admits a Hamiltonian operator  $H : \Psi \rightarrow \Psi$  that allows us to write the laws of evolution as

$$i\hbar \partial_t \psi = H\psi$$

*Justification.*  $\square$

Note: we get time dependent Schroedinger. No basis for time independent.

## X. APPENDIX: MATHEMATICAL PROOFS

In this section we include mathematical demonstration in support of the physical arguments of the paper.

**Theorem X.1.** Let  $X$  and  $Y$  be Hausdorff and second countable topological spaces. Let  $C(X, Y) \equiv \{f : X \rightarrow Y \mid f \text{ is continuous}\}$  be the space of continuous functions endowed with the Compact-Open Topology  $\tau$ . Then  $C(X, Y)$  is a Hausdorff and second countable topological space.

*Proof.* We claim  $C(X, Y)$  is Hausdorff. Let  $f, g \in C(X, Y)$  such that  $f \neq g$ .  $\exists x \in X$  such that  $f(x) \neq g(x)$ . Let  $K = \{x\}$ .  $K$  is compact.  $\square$

**Theorem X.2** (Extended Riesz theorem). Let  $(S, \tau)$  be a locally compact Hausdorff space. Let  $\Lambda = \{\Lambda_U : C(S) \rightarrow \mathbb{R}\}_{U \subseteq S}$  a family of positive linear functionals

such that  $\forall U \subseteq S \ \Lambda_U = \Lambda_{\text{int}(U)}$  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  $\mu$  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  $\mu_U$  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  $\Lambda_U : C(S) \rightarrow \mathbb{R}$ . For the Riesz representation theorem for linear functionals, there exists a unique Borel measure  $\mu_U$  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  $\mu$  such that  $\Lambda_U(\rho) = \int_U \rho d\mu$ . Let  $U \subseteq V \subseteq S$  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  $\Lambda_V(\rho) = \Lambda_U(\rho) + \Lambda_{V \setminus U}(\rho) = \int_U \rho d\mu_U + \int_{V \setminus U} \rho d\mu_{V \setminus U}$ . Therefore  $\int_U \rho d\mu_U = \int_U \rho d\mu_V$ .  $d\mu$  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  $U_\alpha$  is compact.

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

The measure is unique on the whole space.  $\square$

**Theorem X.3.** Let  $\mathcal{L}(S, \mu) \equiv \{\rho : S \rightarrow \mathbb{R} \mid |\int_U \rho d\mu| < \infty \ \forall U \subseteq S\}$ .  $\mathcal{L}(S, \mu) = L^1(S, \mu)$  where  $L^1(S, \mu) = \{\rho : S \rightarrow \mathbb{R} \mid \int_S |\rho| d\mu < \infty\}$

*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  $L^1(S, \mu) \supseteq \mathcal{L}(S, \mu)$ .  $\square$

**Proposition X.4.** Let  $(M, \omega)$  a symplectic manifold. Let  $f : (M, \omega) \rightarrow (M, \omega)$  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  $H : M \rightarrow \mathbb{R}$  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  $v, w \in T_m M$

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  $v'^a = \partial_b \xi(t + dt)^a v^b = (\delta_b^a + \partial_b S^a dt) v^b$ . We have:

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

$v^c w^b \partial_c S_b - v^a w^d \partial_d S_a = 0$  where  $S_b \equiv S^a \omega_{ab}$ . The relationship must true for and pair of vector, therefore  $\partial_a S_b - \partial_b S_a = \text{curl}(S_a) = 0$ .  $S$  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.  $\square$

**Theorem X.5.** *Let  $V$  be a complex vector space. Let  $|\cdot|^2 : V \rightarrow \mathbb{R}$  such that  $|x|^2 \geq 0 \forall x \in V$  and  $|x|^2 = 0 \Leftrightarrow x = 0$ . Let  $x \in V$  and  $P_x : V \rightarrow V$  be a family of projections (i.e.  $P_x$  linear and  $P_x \circ P_x = P_x$ ) such that:  $P_x(y) = y \Leftrightarrow y = ax$  for some  $a \in \mathbb{C}$ ;  $|y|^2 |P_x(y)|^2 = |x|^2 |P_y(x)|^2$ ;  $P_x(P_y(x)) = rx$  for some  $r \in \mathbb{R}$ . Then  $V$  is an inner product space with the product  $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$  defined by  $\langle x, y \rangle x = |x|^2 P_x(y)$ .*

*Proof.* We claim  $\langle \cdot, \cdot \rangle$  is well defined. Let  $x, y \in V$ .  $P_x(P_x(y)) = P_x(y)$  as  $P_x$  is a projection.  $P_x(y) = ax$  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  $x \in V$ .  $\langle x, x \rangle x = |x|^2 P_x(x) = |x|^2 x$ . 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  $a, b \in \mathbb{C}$ .  $\langle x, ay + bz \rangle x = |x|^2 P_x(ay + bz) = a|x|^2 P_x(y) + b|x|^2 P_x(z) = a\langle x, y \rangle x + b\langle x, z \rangle x$  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  $x, y \in V$ .  $|x|^2 |P_x(y)|^2 = |x|^4 |P_x(y)|^2 = |\langle x, y \rangle|^2 |x|^2$ . Using the second property of the projection we have  $|\langle x, y \rangle|^2 = |x|^2 |P_x(y)|^2 = |y|^2 |P_y(x)|^2 = |\langle y, x \rangle|^2$ . The modulus of the product is symmetric as a function of the arguments.

Now consider  $|x|^2 |y|^2 P_x(P_y(x)) = |x|^2 \langle y, x \rangle P_x(y) = \langle y, x \rangle \langle x, y \rangle x$ . 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  $\pi$ .  $\pi$  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.  $\langle x, y \rangle = \langle y, x \rangle^\dagger$   $\square$

**Theorem X.6.** *Let  $C^1(\mathcal{Q}, \mathbb{C})$  be the space of differentiable complex functions defined over  $\mathcal{Q}$ . Let  $q : \mathcal{Q} \rightarrow \mathbb{R}$  a differentiable function. Let  $K_q$  be a covariant linear operator, that is given an invertible differentiable function  $\hat{q} : \mathbb{R} \rightarrow \mathbb{R}$   $K_{\hat{q}} = \partial_{\hat{q}} q K_q$ . 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  $I : \mathbb{R} \rightarrow \mathbb{R}$  be the identify function, that is  $I \circ f = f \forall f : \mathbb{R} \rightarrow \mathbb{R}$ .  $K_q(q) = K_q(I(q)) = \partial_q I(q) K_I(I) = K_I(I)$  by covariance. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  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  $f : \mathbb{R} \rightarrow \mathbb{R}$  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  $K_q(h(q)) = g K_q(f(q)) + f K_q(g(q))$ .  $K_q$  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  $a \in \mathbb{C}$ .  $\square$

**Theorem X.7.** *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  $K = ir \partial_q$  with  $r \in \mathbb{R}$ .*

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

- 
- [1] Shannon, C. E., "A mathematical theory of communication", The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, (1948).
  - [2] Jaynes, E. T., "Information theory and statistical mechanics", Statistical Physics 3, pp. 181–218, (1963).
  - [3] J. V. José, E. J. Saletan, "Classical Dynamics", Cambridge University Press, (1998).

- [4] Gromov, M. L., "Pseudo holomorphic curves in symplectic manifolds", Inventiones Mathematicae 82, pp. 307–347, (1985).
- [5] de Gosson, M. A., "The symplectic camel and the uncertainty principle: the tip of an iceberg?", Foundations of Physics 39, pp. 194–214, (2009).

- [6] Stewart, I., “The symplectic camel”, *Nature* 329, pp. 17–18, (1987).
- [7] Lanczos, C., “The variational principles of mechanics”, University of Toronto Press, (1949).
- [8] Synge, J. L., “Classical dynamics”, *Encyclopedia of Physics* Vol 3/1, Springer (1960).
- [9] 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).