

# 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: August 1, 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 - Mostly done. Needs review.  
Section VI - Rough draft

### 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.

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.

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.

#### III.A. States and their evolution

We start by fixing a *physical system*, meaning something we can interact with and perform measurements

---

\* carcassi@umich.edu

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:

---

<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?

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

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

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

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

stated 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 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

<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.

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 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 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.

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

<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.

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>

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*.

<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?

<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.

*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  $l = s$  is available; for  $n = 2$  we can choose  $l = s$  or  $l = p$ . 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  $\exists s | q_1(s) = l_1, q_2(s) = l_2 \forall l_1 \in q_1(U), l_2 \in q_2(U)$ .

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 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.  $\square$

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

<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.

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 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 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 (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.

<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 interesting in measuring the asymmetry of a particular state without measuring (or without being able to measure) the full state itself.



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 to be a manifold consistently with IV.7, 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.  $\mathcal{C}$  is 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 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.  $\mathcal{C}$  is isomorphic to  $\varrho(\mathcal{C}) \subseteq C(\mathcal{S})$  as a vector space.  $\square$

## 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  $\Lambda_U : \mathcal{C} \rightarrow \mathbb{R}$ .

<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 underline 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).

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  $\Lambda_U = \Lambda_{\text{int}(U)}$ . 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  $c \in \mathcal{C}$  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.1 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.2. 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.  $\square$

#### 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:  $\rho_c(\hat{q}) = \rho(q)_c dq/d\hat{q}$ . 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 differential, 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 differentiable 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 differen-

<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.

tial 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 as 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 differential 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 differential 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  $\hat{q}^j$  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  $|dq^i/dq^j|$  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  $\delta_i^j$  is the Kronecker delta. Let  $q^j$  be another set of state variables.  $q^1(q^j)$  is differentiable as the manifold is differentiable.  $\rho_c(q^j) = q^1(q^j)$  is differentiable over  $U$ .

We claim that the distributions associated with composite states are differentiable. Let  $\rho_c$  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 distri-

bution 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 integral are finite as  $\rho_c, \rho \in L^1(\mathcal{S}, \mu)$ . Both integral 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 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 match the norm given by the vector product defined as  $\langle f, g \rangle = \int_{\mathcal{S}} fg d\mu$ . Moreover  $\mathcal{C}$ , since it excludes discontinuous functions, can never be a complete normed space: it will not include the limit for all Cauchy sequences, it cannot be a Banach or a Hilbert space. Such construction, though, is still

<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.

useful. Consider the following expressions:

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

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 the Jacobian has to exist but 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  $|\partial_i \hat{k}^j| = 1/d_q \hat{q}$ . This only happens if  $\partial_i \hat{k}^j$  is a  $1 \times 1$  matrix: there is only one  $k$ . This gives  $\partial_k \hat{k} = d_q \hat{q}$  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 able to write the volume of integration in terms for 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  $\omega$  given by the

<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.

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  $\omega = \Sigma dq^i \wedge dp_i$  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 unit transformation. Let  $(q, k^i)$  with  $i = 1 \dots n-1$  be a set of state variable. The Jacobian determinant  $|J| = d_q \hat{q} |\partial_i \hat{k}^j| = 1$  as densities must be invariant.  $|\partial_i \hat{k}^j| = d_{\hat{q}} q$ .  $\partial_i \hat{k}^j$  is one dimensional as its components must be fully specified by the previous relationship. Let  $k = k^1$ .  $\partial_k \hat{k} = d_{\hat{q}} q$ .  $\hat{k}(q, 0) = 0$  as a change in units does not change the 0 value.  $\hat{k} = d_{\hat{q}} q k$ .  $k$  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 differential 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 differential manifold  $\mathcal{Q}$  and  $\omega$  is the canonical two-form. Let  $q^i$  be a set of  $n$  continuous independent state-variable 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. Each degree of freedom must define state-variable-distribution independently. 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  $\int_{U \subseteq \mathcal{S}} \rho_c \omega$  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.  $\omega = \Sigma dq^i \wedge dp_i$ .  $\mathcal{S} = (T^*\mathcal{Q}, \omega)$  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 finite amount of material in a finite number of states.

In other words: this is not a case were a difference at small scale has a small difference at large scale. 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. Or more precisely: that the processes we use to distinguish those quantities cannot identify single instances, as that is all we can say experimentally. The quantum case does not change this, as the use of densities (in the form of the wave function) and conjugate quantities are even more prominent.

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>15</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.

Consistently with what we said in Assumption I, if we defined a state for each particle, then a deterministic and reversible evolution on such 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 more 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 pool ball we can imagine marking one spot with a red marker. This allows us to track it as the ball moves or collides with other balls. For an electron, instead, we do not have a way to mark a piece. In fact when two electrons scatter we can't even tell which is which, let alone what portion went where. The classical assumption may not hold because we do not have suitable physical processes at our disposal.

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

The second, more conceptual, problem is that the idealizations that allow us to hold Assumption I break down as the parts get smaller. One such idealization is when the system under study is "big enough". Recall the cannonball whose motion is sufficiently unaffected by the photons that scatter off its surface. Clearly this breaks down for infinitesimal parts: as we keep dividing the material, we will find an amount that will be affected by the interaction with the environment, and therefore not sufficiently isolated. The other idealization is when the interaction between system and environment is not symmetrical. For example, suppose the electric charge of the system under study is much larger than the charges in the environment. The electromagnetic field generated can be considered determined only by our system with good approximation. But this also breaks down: if each part of the environment is made of the same homogeneous material the charge density will be the same and the interaction between particles will be symmetrical.

In other words: the definition of a state requires some kind of asymmetry between system under study and environment. The environment should not affect the system, so that we can define an independent state, yet the system must affect the environment, so that we have external process that allows to define physical distinguishability. If we try to assume that both the system and environment are really made of infinitesimal homogeneous parts, then the claim to that asymmetry is lost. Therefore our justification for holding the assumption has to be that the pieces are small compared to some scale (they can be regarded as infinitesimal) but large compared to another (they are little affected by the environment). For example, water is a homogeneous material if we consider it made up of parts sufficiently small for the scale of the problem at hand but big enough to contain a large number of molecules such that the average motion is sufficiently stable.

Other conceptual problems stem from the following two competing requirements: on one side we want the system under study to be isolated so that we can define independent states; on the other side we want the envi-

<sup>15</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.

ronment to not be isolated so that we can define physical processes that interact with the system. The problem is that the same physical object could be either under study or part of the environment depending on the situation. In other words: we want systems to be interacting so that they can be physically aware of each other, yet we want them not interacting so that we can properly define states on them. We clearly can't have it both ways. 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 cannot work.

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 we have pieces with no unstated part that are physically aware of each others. But this does not actually solve the problem of independent oblivious components. Let's ignore the question of how we were able to define the states of the parts. Let's also ignore that the type of forces one would typically use (i.e. ones that conserve energy or ones with scalar and vector potentials) are precisely the ones we derive by assuming the pieces are not interacting with each other. The actual 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>16</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 the particles are in principle interacting, the system is described by degrees of freedom that evolve independently, oblivious to each other.

The same type of problem arises on the whole system when we consider its interaction with our probe or other parts of the environment. Say  $a \in A$  is the state of our system and  $b \in B$  the state of our probe, the future state  $\hat{a} = f(a)$  because the system is deterministic, but  $\hat{b} = g(a, b)$  because the probe needs to be affected by the system. This means  $(a, b) \mapsto (\hat{a}, \hat{b})$  is not a reversible map. In other words: we can't expect to have deterministic and reversible evolution at all levels of decomposable systems if the parts interact in any physically meaningful way.

The moral is that we should *always* beware of thinking that the state of a system is its full description: there is always some unstated part that allows continuous interaction between the system and the environment. That is the part that cannot be isolated (i.e. we can't take the system out of the physical universe) and therefore can't be assigned a state.

One final conceptual problem is with time itself: a deterministic and reversible system 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.

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 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

<sup>16</sup> Also note that any Hamiltonian is locally isomorphic to a free particle (citation needed)

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. 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  $H : T^*\mathcal{Q} \rightarrow \mathbb{R}$  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  $H$  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.3 the infinitesimal displacement  $S$  admits a potential  $H$  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. 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 straight forward.

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 evolution of an invariant density distribution, and not the description of a different type of 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 are defined



at equal time and 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. Note that this does not affect our definition of determinism and reversibility: we still can tell final state from initial state and vice-versa. We just need to specify the time as well. At this point we have no way to do that. For example, the Hamiltonian  $H : T^*\mathcal{Q} \rightarrow \mathbb{R}$  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  $d_t \hat{t} d_{\hat{t}} q^i = \partial_{p_i} H$ . As both  $q^i$  and  $H$  are independent of time, they 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.

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 is common in physics and engineering, which is why we kept the standard name, 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  $\mathcal{S}_{t=t_0}$  that includes all

possible configuration at a particular time. Such hyper-surface will need to cut across all evolutions once and only once, therefore not all functions of  $\bar{\mathcal{S}}$  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 an  $2n+1$  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 an  $2n+1$  dimensional manifold.  $T^*\mathcal{Q}$  is a differentiable manifold. Time evolution is a diffeomorphism by V.7 therefore derivatives between state variables and time variables always exist and are invertible.  $\bar{\mathcal{S}}$  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 an hyper-surface.  $\square$

For the composite state we need to take a different approach. A distribution of material over states at equal time  $\rho_c : \mathcal{S}_{t=t_0} \rightarrow \mathbb{R}$  will depend on the choice of time variable, therefore the complete state space will not be useful as it will not be variable independent. What we are really interested in is the *evolution space* of a classical material, which we'll indicated 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, but 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  $\bar{\mathcal{C}} \cong G \subseteq C^1(\bar{\mathcal{S}}) \cap L_t^1(\bar{\mathcal{S}})$ . Let  $\rho_{\bar{c}}$  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.  $\rho_{\bar{c}} \in C^1(\bar{\mathcal{S}})$

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 space. Note that the transported state variables are perfect to label the particle state evolution evolutions  $\bar{\mathcal{Q}} \subset \bar{\mathcal{S}}$ : 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  $\bar{\mathcal{Q}}$ . To identify the full evolution, then, we just need to specify the distribution only on a region that cuts across all particle evolution, 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  $t$ . In fact, the idea that the set of units to identify a physical configuration at a particular time did not include time should look, retrospectively, shaky at best. 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 is valid). As we said before, the time variable is not simply another state variable: it doesn't identify extra states 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 each instant in time. In this sense, time is not just like another state variable.<sup>17</sup>

As in the time independent case, we are interested in invariant densities which are defined on co-vectors. The co-vector  $k_i dq^i$  is completed with a time component  $\bar{\omega} dt$ .<sup>18</sup> As state variable  $q^i$  have 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 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 states at a particular time, 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 set  $\bar{\mathcal{S}}$  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  $dt \wedge dE$ .

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 count the possibilities within a degree of freedom, adjusting the metric on  $T^*\mathcal{Q}$  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

<sup>17</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.

<sup>18</sup>  $\bar{\omega}$  is the classical analogue of the wave frequency. We use  $\bar{\omega}$  to distinguish from the phase-space metric  $\omega$ .

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 that fact that the vector product is conserved within a degree of freedom is equivalent to saying that changing variables conserves the possibilities within a d.o.f. In fact, the flow of an infinitesimal transformation that preserves the vector product on a surface is a divergence free field. The conservation of the Minkowsky 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 flow of an infinitesimal transformation that preserves the Minkowsky product on a surface is a curl free field.

We can capture the above discussion by stating that the complete state space  $\hat{\mathcal{S}}$  is a hypersurface of the symplectic manifold  $(T^*\mathcal{M}, \omega)$ , where  $\omega \equiv \sum dq^i \wedge dp_i - dt \wedge dE$ . This is the manifold that allows us to define time-and-state-variable-invariant densities. The metric allows us to measure the cardinality of possibilities.

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

*Justification.* We claim the space of physical objects that allows a time-state-invariant-variable 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  $E = \hbar\bar{\omega}$ . Let  $\omega$  be the metric. The terms across different d.o.f. must be zero as they are not invariant under unit transformations.  $\omega = \sum dq^i \wedge dp_i \pm dt \wedge dE$ . 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  $\hat{\mathcal{S}}$  is a hyper-surface of  $(T^*\mathcal{M}, \omega)$ .  $\hat{\mathcal{S}} \subseteq (T^*\mathcal{M}, \omega)$  as states are physical objects that allow time-state-invariant-variable densities. Consider the map  $f : T^*\mathcal{M} \rightarrow \hat{\mathcal{S}} \mid (q^i, p_i, t, E) \mapsto (q^i, p_i, t)$ .  $f$  is an open map. The topology of  $\hat{\mathcal{S}}$  is the subspace topology. Let  $\rho : T^*\mathcal{M} \rightarrow \mathbb{R}$  be an invariant density.  $\rho$  is differentiable by V.4. It's restriction on  $\hat{\mathcal{S}}$  is also an invariant distribution and therefore differentiable. The inclusion map  $\hat{\mathcal{S}} \hookrightarrow T^*\mathcal{M}$  is a smooth embedding.  $\hat{\mathcal{S}}$  is an embedded submanifold of co-dimension 1.  $\hat{\mathcal{S}}$  is an 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:  $\mathcal{J}_{\Delta s} : T^*\mathcal{M} \rightarrow T^*\mathcal{M}$  such that  $\mathcal{J}_{\Delta s}^*\omega = \omega$  where  $\mathcal{J}_{\Delta s}^*$  is the pullback of  $\mathcal{J}_{\Delta s}$ .

*Justification.* We claim  $\mathcal{J}_{\Delta s}$  is a self-homeomorphism on  $T^*\mathcal{M}$ . The extended state space for particles of a homogeneous classical material is  $T^*\mathcal{M}$  by VI.3.  $\mathcal{J}_{\Delta s}$  is a deterministic and reversible evolution map and by IV.3 is a self-homeomorphism.

We claim  $\mathcal{J}_{\Delta s}$  is a symplectomorphism on  $(T^*\mathcal{M}, \omega)$ . The distribution on final states must be defined. The Jacobian for  $\mathcal{J}_{\Delta s}$  exists and is non-zero.  $\mathcal{J}_{\Delta s}$  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.  $\mathcal{J}_{\Delta s}$  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 extended 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  $S^a = d\xi^a/ds$  admits a potential  $\mathcal{H}$  such that  $S^a \omega_{ab} = \partial_b \mathcal{H}$ . The extended state space  $T^*\mathcal{M}$  of a classical homogeneous material is a symplectic manifold by VI.3. The map for infinitesimal evolution  $\mathcal{J}_{ds}$  is an infinitesimal self-symplectomorphism by VI.4. By X.3 the infinitesimal displacement  $S$  admits a potential  $\mathcal{H}$  such that  $S^a \omega_{ab} = \partial_b \mathcal{H}$ .

We claim the state variables evolve according to the extended Hamilton's 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>19</sup>

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

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

We claim  $d_s t$  partitions the space. Let  $s_1, s_2 \in T^*\mathcal{M}$  two physical states connected by deterministic and reversible evolution.  $\text{sign}(d_s t(s_1)) = \text{sign}(d_s t(s_2))$ . Let  $U \equiv \{s \in T^*\mathcal{M} \mid d_s t(s) > 0\}$  and  $V \equiv \{s \in T^*\mathcal{M} \mid d_s t(s) < 0\}$ .  $U \cap V = \emptyset$ . Let  $\gamma : [s_0, s_1] \rightarrow T^*\mathcal{M}$  the trajectory given by deterministic and reversible evolution, where  $[s_0, s_1]$  is the range of the parametrization. Either  $\gamma([s_0, s_1]) \subseteq U$  or  $\gamma([s_0, s_1]) \subseteq V$ .  $\square$

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

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 extended Hamiltonian is  $\mathcal{H} = \lambda(t, E, q^i, p_i) (H(t, q^i, p_i) - E)$  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  $\lambda : T^*\mathcal{M} \rightarrow \mathbb{R}$  where  $\lambda(\xi^a) = d_s t$ .  $\bar{\mathcal{S}}$  is the level set for  $\mathcal{H} = 0$ .

*Justification.* We claim the form of the extended Hamiltonian is  $\mathcal{H} = \lambda(t, E, q^i, p_i) (H(t, q^i, p_i) - E)$ . Let  $\mathcal{H}$  the extended Hamiltonian and  $H$  be Hamiltonian for a particular choice of time and state variables. We have  $d_s q^i = \partial_{p_i} \mathcal{H} = d_s t d_t q^i = -\lambda \partial_{p_i} H$  where  $\lambda : T^*\mathcal{M} \rightarrow \mathbb{R} \mid \lambda \mapsto d_s t$ .  $d_s p_i = -\partial_{q^i} \mathcal{H} = d_s t d_t p_i = \lambda \partial_{q^i} H$ . As  $\lambda \neq 0$  for all physical states, we can set  $\mathcal{H} = \lambda(t, E, q^i, p_i) (H(t, q^i, p_i) - E + h(t, E, q^i, p_i))$  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.

We claim  $E = H(t, q^i, p_i)$  for physical states.  $\lambda = d_s t = -\partial_E \mathcal{H} = -\partial_E \lambda (H - E + h) - \lambda (-1)$  for any choice of  $t$ .  $H - E = 0$ .

We claim  $\mathcal{H} = 0$  for physical states.  $\mathcal{H} = \lambda(H - E) = 0$ .  $\square$

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

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

In the next section we'll derive 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.

<sup>19</sup> This represents the classical analogue for quantum anti-particle states.

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

**Proposition VI.9.** Let  $\rho_{\bar{c}}(s)$  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  $d_s\rho_c(s) = \{\rho_c(s), \mathcal{H}\}$ . Let  $\rho_c(s)$  be differentiable.  $d_s\rho_c(s) = \partial_a\rho_c(s)d_s\xi^a = \partial_a\rho_c(s)\omega^{ab}\partial_b\mathcal{H} = \{\rho_c(s), \mathcal{H}\}$ .

We claim  $\mathcal{H}\rho_{\bar{c}} = 0 \forall \bar{c} \in \bar{\mathcal{C}}$ . Let  $\rho_{\bar{c}} : T^*\mathcal{M} \rightarrow \mathbb{R}$  be the distribution associated with an evolution  $\bar{c}$ .  $\text{supp}(\rho_{\bar{c}}) \subseteq \bar{\mathcal{S}}$  as the distribution is defined on the complete state space of the particles.  $\text{supp}(\mathcal{H}) = T^*\mathcal{M} \setminus \bar{\mathcal{S}}$  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  $\gamma : \mathbb{R} \rightarrow T^*\mathcal{M}$  be a parametrization for a particle evolution  $\bar{\gamma}$ . Then  $\rho_c(\lambda(s)) = \rho_c(\lambda(s + \Delta s))$ .  $d_s\rho_c = 0$ .  $\square$

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 chapter 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 of invariant densities. It would not have arisen with the use of a discrete topology.

We have also seen that anti-particle states, a feature generally associated with quantum field theory, is 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 extended Hamiltonian in VI.8 is quadratic in  $E$ , which can have both positive and negative solutions on 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 the extended Hamiltonian in VI.8 with the usual formal substitution, we have the Klein-Gordon equation using space-like convention rescaled by a factor for  $2m$ . As we'll see later, to this parallel in the mathematical formalism corresponds a parallel in the physics description.

As a final thought, we note how the physics maps less elegantly to the math in the time dependent case. The function space for  $\bar{\mathcal{C}}$  is not simply the Lebesgue integrable

differentiable functions, but those that are integrable on particular hyper-surfaces. The complete state space  $\bar{\mathcal{S}}$  is a hyper-surface of  $T^*\mathcal{M}$ . The extended 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, label those (instead of the states). The distribution of material would simply be a distribution over the particle evolutions, a Lebesgue integrable differentiable functions.

## VII. ELEMENTARY SYSTEMS

Reductionism by itself would not help us describing a system if the state and the evolution of the pieces were more complicated to study. In this section we further focus our attention to those system that not only are infinitesimally reducible, but their internal dynamics can be ignored. That is, we will assume that the motion in space is necessary and sufficient to reconstruct the state and its dynamics. What we'll find is that the dynamics of each infinitesimal part follows classical Lagrangian mechanics. Furthermore, the Hamiltonian of the system is constrained to the one of a free particles 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 laws of evolution, 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 and the Lagrangian.

In this work we make the link conceptually explicit by assuming the following:

**Assumption III** (Kinematic equivalence). *Studying the kinematic of the system (i.e. trajectories in space) is equivalent to study the dynamics of the system (i.e. state trajectories in state space).*

*Rationale.* The idea is that the system internal dynamics has no affect on the 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 trajectory. We can assume kinematic equivalence. Consider the motion of a helicopter under gravitational and inertial forces: the pilot and the amount of fuel in the tank do affect the trajectory. We cannot assume kinematic equivalence.

The assumption clearly cannot be taken in general since many systems do not posses this trait. Yet, as this assumption allows to recover the standard link between kinematic and dynamical quantities, there is something

fundamental about it. That is: if we assume that the system has no discernible internal dynamics, then the only dynamics that needs to be described is the one associated with its trajectory in space: kinematic equivalence holds.

This is in fact consistent with observation, though elementary particles are quantum fields therefore some caveats are in order. Massless particles should be excluded as they cannot be localized, which makes it hard to talk about a proper trajectory. Spin is not associated to kinematic quantities: nothing is thought to be actually spinning. As such we can regard it as the internal dynamic of the system. A massive spin 0 system is one that can be thought as localized and as having no internal structure. For such system we would expect kinematic equivalence to hold and it does: from kinematic momentum we can reconstruct conjugate momentum and vice-versa.

Therefore, with the appropriate caveat that we are talking about a quantum particle which we haven't yet introduced, the simplest systems are the ones that just describe a trajectory in space. This raises the opposite question: why all systems have *at least* to describe a trajectory in space? This seems motivated by the previous assumption of reducibility: spatial extent is the one way that all materials are reduced. Therefore all particles of all materials, regardless of whether they are elementary or not, are will include spatial degrees of freedom.

It should be also 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 dynamics must be identified scalar, vector or tensor quantities. Spin is, in some sense, a quantized version of those.

While we can make an argument why kinematic equivalence is important because it is valid for elementary system, we cannot actually make an argument why nature should be reducible to elementary system. While it fits our physical intuition, we have no a priori argument for it. We can imagine, in fact, that as we reduce the behavior of one material to its constituents we find the parts to possess richer and richer states. For example, most macroscopic object are neutral though the parts are charged. We could imagine that, while dividing into parts, we find more and more types of charges and forces. We do not have an argument why this cannot happen.  $\square$

## VIII. PRELUDE TO QUANTUM

TODO

## VIII.A. Classical uncertainty principle

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

$$\sigma_q \sigma_p \geq \exp(H_0 - 1)/2\pi$$

## IX. QUANTUM SYSTEMS

TODO

## X. APPENDIX: MATHEMATICAL PROOFS

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

**Theorem X.1** (Extended Riesz theorem). *Let  $(S, \tau)$  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  $\rho \in C(S)$ .  $\Lambda_V(\rho) = \int_V \rho d\mu_V = \int_U \rho d\mu_U + \int_{V \setminus U} \rho d\mu_V$ . Also  $\Lambda_V(\rho) = \Lambda_U(\rho) + \Lambda_{V \setminus U}(\rho) = \int_U \rho d\mu_U + \int_{V \setminus U} \rho d\mu_{\overline{V \setminus U}}$ . Therefore  $\int_U \rho d\mu_U = \int_U \rho d\mu_V$ .  $d\mu$  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.

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

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

**Theorem X.2.** 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.3.** 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 exist a function  $H : M \rightarrow \mathbb{R}$  such that  $S^a \omega_{ab} = \partial_b H$ .

*Justification.* 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  $v^a, w^b$  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 any 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  $S_a = \partial_a H$ .  $\square$

- 
- [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).