ASSUMPTION OF EQUAL APRIORI PROPABILITIES: CLASSICAL CASE

K.M. HALPERN

Contents

1. Intro 1
2. Brief review of some topology and differential geometry 4
2.1. Some useful definitions and results in topology 4
2.2. Note on different topologies on the same set 6
2.3. Some useful definitions and results for manifolds 8
3. The sets of microstates: Ω and Ω_E. 14
3.1. Ω and Ω_E as topological spaces. 15
3.2. Properties of Ω_E. 16
4. Review of some relevant measure theory 17
4.1. Basics 17
4.2. Borel algebras 18
4.3. Borel and Radon measures 20
4.4. Completion of a measure 21
5. Uniform probabilities 24
5.1. Uniform probability field 24
5.2. Aside: Uniform random variable 25
6. Measure induced by a volume form on a manifold 28
6.1. Premeasure as a functional 29
6.2. Riesz representation theorem 30
7. Review of symplectic manifolds 35
7.1. Brief review of Inner Products 36
7.2. Symplectic forms on Vector Spaces 38
7.3. Symplectic manifolds 38
7.4. T^*Q as a symplectic manifold 40
7.5. Tautological 1-form on T^*Q 40
7.6. Tautological 2-form on T^*Q 42
8. Putting it all together 43
References 45

1. Intro

As a student, more years ago than I care to count, I was bothered by something we were taught in statistical mechanics. Actually, I was bothered by a *lot* of what we were taught in stat mech — but I'm going to focus on one particular issue in this post.

If you're a physics student, you almost certainly have encountered something called the "assumption of equal apriori probabilities" — which I'll refer to here as the "AEAP".

This should not to be confused with "AEP", which commonly refers to the "asymptotic equipartition theorem", and is something completely different.

The AEAP says that for the microcanonical ensemble all accessible states have equal probability. I.e., all microstates with a given energy are equally likely. This sounds simple enough, but — as with almost everything in statistical mechanics — beneath its superficial simplicity lie some thorny mathematical issues.

There are two potentially problematic aspects to the AEAP: one philosophical and one technical. The philosophical one is more prosaic (and turns out to be largely immaterial because of asymptotic phenomena), and we won't address it here. The more interesting question is whether the AEAP is even meaningful. Can we impose a uniform probability distribution on the relevant set of microstates?

At heart, the AEAP is nothing more than the usual "Principal of Maximum Entropy". However, there are some subtleties related to the definition of entropy in this context and the physical basis for such an assumption. Symmetries and the fungibility of the components can offer some justiication, as can asymptotic phenomena in the case where the components themselves are aggregates of smaller constituents. It turns out that similar asymptotic phenomena, predominantly in the form of the central limit theorems of statistics, render the details of the AEAP (and to some extent its very existence) less significant to the rest of statistical mechanics than may at first appear to be the case.

We're using the term "uniform probability distribution" informally at this point. Depending on the context, it could involve a set of discrete probabilities, a probability density, or a probability field. Think of it as a casual way of saying "something that corresponds to our notion of uniform." Obviously, we'll need to be precise when we actually discuss it in detail.

The first question is: what do we mean by "uniform"? For a finite set S, this is straightforward enough — we just assign each element a probability $1/|S|$. However, things quickly get murky when we move to infinite sets. For example, it is impossible to impose a uniform probability distribution on the integers.

Boundedness is not the only issue. The continuous case is problematic even when bounded. Consider the unit cube $[0,1]^3$, viewed as a manifold. One approach would be to assign the "obvious" uniform probability density $\rho(x) = 1$. However, doing so assumes a preferred set of coordinates. If we used spherical coordinates instead, our "uniform" density wouldn't look uniform any more.

If we take $[0,1]^3$ to have the usual Euclidean metric, then we do indeed have a preferred class of coordinate systems (the orthonormal ones) and a natural notion of volume that is the same for all of them. However, this is because of the additional structure of the metric. As a mere manifold, $[0,1]^3$ has no natural notion of volume. Since it's orientable, it can admit volume forms — but we have no reason to designate any particular volume form as special.

The dartboard "paradox" encapsulates this issue. Suppose a dartboard is broken into a grid of equal-sized squares. Somebody throws darts at the board and the darts have uniform probability of landing anywhere. What fraction land in each square? One answer is that every x and y value is equally likely for the dart, so each square receives the same number of darts on average. But what if we reason differently? Each position has a radius and an angle relative to some reference direction. The probability that a dart lands at any given radius is uniform and the probability that it lands at any given angle is uniform. So the darts should cluster near the center. One may decry this as obviously wrong, since different radii do *not* have equal probabilities. However, the flaw lies not in this claim but in the statement of the problem. We said "uniform" but did not specify what we meant by it. One assumption is that we meant uniform relative to Cartesian coordinates. Under that assumption, the 2nd approach is patently wrong. However, we could just as well have meant uniform relative to polar coordinates (i.e. viewing the distribution as uniform on the rectangle $(0, R] \times [0, 2\pi)$). In that case, the first approach is wrong for the very same reason. We may argue that an actual physical process would most likely lead to the first distribution, but that is an additional assumption. We could easily concoct a scenario where it yields the second (ex. a machine that randomly determines an angle and radius and throws darts accordingly). The point is that our notion of "uniform" requires a choice. If we wish this choice to be "natural" in the sense that it reflects and is compatible with the physical assumptions of the problem, then we require additional structure. In our case, this structure could come in the form of a preferred coordinate system or a metric or information about how a dart-throwing machine operates.

Once we've decided what we mean by "uniform" probabilities, we'll need to ascertain whether they are possible in a given situation. In the case of the AEAP, we'll have to consider two classes of systems: classical and quantum. For each, we must:

(1) Identify the set of relevant microstates with a given energy and understand its mathematical structure and properties.

(2) Ascertain what it means to have a "uniform distribution" in this context.
(3) Determine what constraints and/or structure is necessary for such a uniform distribution to exist.
(4) Establish how the physics of the system provides such constraints and/or structure, leading to a unique, natural, and meaningful (for the purposes of the AEAP) way to define a "uniform distribution" on the relevant set of microstates.

Step (4) deserves some clarification. It is not enough to say that our set of microstates is a mathematical shmugel and that we need a shmoo on it in order to define a notion of uniform probabilities and that this particular shmugel admits a shmoo. We still may end up with the dartboard problem. To partly preview what is to come, if our physical assumptions tell us that the schmugel is a manifold and the shmoo is a volume form, we not only need the manifold to be orientable (thus admitting volume forms), but we need the physical assumptions to identify a unique "natural" choice of volume form.

More generally, the physical assumptions must either constrain the schmugel to accept only a single (and thus, as the only choice, "natural") shmoo, or they must impose additional structure which induces a "natural" choice of shmoo.

The distinction is largely pedantic. To again offer a preview, if we define the shmugel to be a manifold and the shmoo to be a volume form then we have the second scenario and require that the physics also picks out a preferred volume form. In this case, we have additional information to the effect that our manifold is actually a cotangent bundle and thus has a natural symplectic volume form. On the other hand, if we *start* by defining the shmugel to be a cotangent bundle and the shmoo to be a symplectic volume form, then we essentially have constrained the problem at the outset. In both cases, the information comes from the physical assumptions — and we just choose to classify it in different ways.

The point is that we need to uniquely identify a "special" shmoo without the introduction of nonphysical choices or structures. The physical assumptions alone must provide a canonical mechanism for defining our uniform distribution *and* this uniform distribution must be meaningful for the purposes of the AEAP. I.e., the uniform distribution is not just a technical nicety; it must be the correct definition to plug into the AEAP.

This last point is important. Ultimately, the proof is in the pudding. It is the right choice if it makes statistical mechanics work. If our "canonical" choice does *not* work, we must ask why. In that case, either the AEAP is an incorrect assumption for statistical mechanics, or our fundamental physical assumptions are incomplete or wrong, or there is some other canonical choice which we overlooked. Even if we find such an alternate choice, we still need to understand why the definition of "volume" via our initial choice was inappropriate for the AEAP. Fortunately, this challenging exercise will prove unnecessary. In the classical case, a canonical choice presents itself in the symplectic volume on the phase space and (as far as we can tell) properly fuels the AEAP. Liouville's theorem and other aspects of classical mechanics offer good reason to view it as the correct choice. We won't delve into this in detail.

In the classical case, our aforementioned steps translate into the following.

(1) Establish that the overall set of classical microstates is a cotangent bundle and that the set of microstates whose energy lies in any small open interval of energies is a submanifold of the same dimension.
(2) Discuss how any finite measure on a σ-algebra gives rise to a notion of "uniform" probabilities, simply via normalization into a probability measure.
(3) See how any volume form on a manifold induces a unique compatible, locally-finite, almost-regular measure on the corresponding Borel algebra.
(4) Show how the symplectic structure intrinsic to a cotangent bundle gives us a natural volume form on the overall state space, which then induces a natural volume form on the set of microstates corresponding to a small open interval of energies.

Putting all this together, we'll see that it is the physical assumptions that (i) *our overall state space is a cotangent bundle* and (ii) *our Hamiltonian is continuous* which essentially give rise to the natural notion of uniform probabilities we require for the AEAP. As it happens, we'll ultimately require a 3rd physical assumption as well — but one which does not constrain us in any meaningful way.

As a point of nomenclature, we'll often sloppily speak of the "state space" rather than "set of states", despite not having confirmed that is is either a topological or vector space yet. We do so from laziness rather than pointed anticipation of its true nature. As we will see, it *is* a topological space in both classical and quantum mechanics, so this is somewhat justified. However, if we want to be fastidious we should consider it a mere set until we establish its structure.

It turns out that the quantum case is quite a bit trickier than the classical one. We can follow a similar regimen, but the relevant structures and behaviors aren't as straightforward. For simple finite-state systems (spin systems, qubits, etc), we can just do the obvious: assign each state probability $1/N$. However, things get trickier when the state space is infinite. Although the quantum state space *is* a manifold, it is an infinite-dimensional one in general, raising a host of technical difficulties. For this reason, we will defer a proper treatment to a future post.

Note that this is not merely a matter of discrete vs continuous spectra (which we'll discuss below). The question is the size and nature of the state space. For example, we could have a very large state space that maps to a single energy (ex. if there is a symmetry). A finite spectrum does not imply a finite state space — though the two often go hand in hand. Obviously, an infinite spectrum does require an infinite state space.

Formally, the quantum mechanical state space is a projective Hilbert space. In the finite dimensional case, this is CP^n (complex projective space), which is a compact Kahler manifold. It is orientable and has a Riemannian metric, a symplectic form, and a distance metric (all inherited from the Hermitian inner product on $\mathbb{C}^{n+1}$). We may be tempted to replicate the classical approach: deriving a symplectic volume form, etc. In fact, CP^n is a lot nicer than the classical state space in many ways, since it is compact. However, the Hamiltonian no longer is a function from the state space to $\mathbb{R}$. It is an operator on the relevant Hilbert space $\mathbb{C}^{n+1}$. Our Ω_E (to be defined shortly in the classical case) becomes less trivial to define. The Hamiltonian only specifies a fixed energy value for energy eigenstates, not general pure states. Since Ω_E (essentially) involves states of a fixed E, this particular problem is mitigated. However, there are related hurdles which arise, and we still must be very careful. A bigger problem is that almost all quantum systems of interest involve an infinite-dimensional projective space. Proving that the necessary properties and results hold for such a manifold is nontrivial. For these reasons, the subject deserves its own treatment.

For similar reasons, we'll focus here on classical systems with a finite number of degrees of freedom — and thus a finite-dimensional configuration manifold. When we refer to a "classical system", we'll mean subject to this constraint.

This includes all the usual point-particle and rigid-body systems — as well as classical fields (ex. fluids) that are constrained (bounded, regularized, etc) in a way that keeps their configuration manifolds finite-dimensional. Many of the results extend to the infinite-dimensional case, but we can't take this for granted.

We'll start with a brief review of point-set topology and manifolds, discuss the relevant set of microstates for the AEAP, review some measure theory, consider the meaning of "uniform" in the context of a general probability measure, discuss how a volume form on a manifold induces a natural measure on its Borel algebra, review some symplectic geometry, and finally bring it all together.

A basic knowledge of topology, differential geometry, and measure theory is assumed, but these notes are otherwise meant to be self-contained. The impatient reader may wish to skip to section 8, which describes the entire regimen in detail, and then peruse the other sections as needed.

2. Brief review of some topology and differential geometry

2.1. **Some useful definitions and results in topology.**

A topological space is:

- **Compact** if every open cover has a finite subcover. A subset is compact if it is compact in the subspace topology.
- **Locally compact** if every point has a compact neighborhood (i.e. every point is contained in an open neighborhood which is contained in a compact set).
- **Paracompact** if every open cover has a locally finite open refinement.

  Let $\{U_i\}$ be a cover of S. A "refinement of cover" $U = \{U_i\}$ is a cover V=$\{V_i\}$ of S s.t. each $V_i \subseteq U_j$ for some U_j. Note that (unlike a refinement of a partition), this does *not* require that each U_j be filled in by the V_i's. The $\{V_i\}$ do cover U_j (since they cover S), but it is quite possible they do not fill it. Ex. let $S = \{1,2,3,4\}$ and let $U = \{(1,2,3,4),(2,3)\}$ and $V = \{(1,3,4),(2)\}$. Then V is a refinement of U, even though we can't write $(2,3)$ as a union of V_i's. An open cover is "locally finite" if for each $x \in S$, there exists some open set containing x which intersects only finitely many of the V's.
- **Hausdorff** if any two points have disjoint open neighborhoods.
- **Second-Countable** if it has a countable basis. I.e., any open set is a union (possibly infinite) of members of this countable basis.
- **Connected** if it isn't the union of two disjoint open sets.

The following relationships hold:

- (i) Compact implies paracompact and locally compact.
- (ii) Second-countable, Hausdorff, and locally compact together imply paracompact.
- (iii) Neither paracompact nor locally compact implies the other.

  (i) follows trivially from the definitions. A good proof of (ii) can be found in [1], theorem 2.6. For (iii), [2] provides a number of counterexamples in both directions (these can be located using the "general reference chart" at the end of the book).

A function $f : X \to Y$ between topological spaces is **continuous** (relative to the specific topologies on X and Y) if the inverse image of every open set is open (or, equivalently, the inverse image of every closed set is closed).

Given a function $f : X \to V$ from a topological space X to a vector space V (or to any algebraic structure with the notion of an identity element), the **support** of f is $f^{-1}(X - \{0\})$. I.e., all points that don't map to 0. The **closed support** of f is the closure of $f^{-1}(X - \{0\})$ (i.e. the intersection of all closed sets containing $f^{-1}(X - \{0\})$). Finally, f is said to have **compact support** if its closed support is a compact subset of X. We say that a set is **connected in** X if it is a connected space in the subspace topology.

Prop 2.1: Let $f : X \to Y$ be a continuous function between topological spaces. Then:

- (i) If O' is open in Y, $f^{-1}(O')$ is open in X.
- (ii) If C' is closed in Y, $f^{-1}(C')$ is closed in X.
- (iii) If X is compact and Y is Hausdorff and K' is compact in Y, then $f^{-1}(K')$ is compact in X.
- (iv) If K is compact in X, then $f(K)$ is compact in Y.
- (v) If S is connected in X, then $f(S)$ is connected in Y.

  Pf: (i-iii): (i) is the definition, and (ii) follows because $\overline{f^{-1}(s)} = f^{-1}(\overline{s})$. (iii) In a Hausdorff space, any compact set is closed (see [3], theorem 26.3). The inverse image of a closed set is closed under a continuous map, so the inverse image of a compact set is closed in X. However, any closed subset of a compact space is compact (see [3], theorem 26.2).

Pf: (iv) Let $U \subset X$ be a compact set. Consider $f(U)$. To be compact, every open cover must have a finite subcover. Let $\{O_i\}$ be an open cover of $f(U)$ (with the $O_i \subset Y$, obviously). As the inverse image of an open set, $f^{-1}(\cup O_i)$ is open in X. Moreover, $U \subseteq f^{-1}(f(U)) \subseteq f^{-1}(\cup O_i)$ (with the 2nd part since $\cup O_i$ is a cover of $f(U)$), so $U \subseteq f^{-1}(\cup O_i)$. However, $f^{-1}(\cup O_i) = \cup f^{-1}(O_i)$. I.e., $U \subseteq \cup f^{-1}(O_i)$. Each of the latter is open, so they form an open cover of U in X. Since U is compact, this open cover must have a finite subcover. Designate this $\{U_i\}$ (i.e. $\{U_i\} \subseteq \{f^{-1}(O_i)\}$, with equality impossible if the latter is infinite). This means $U \subset \cup U_i$, where the right side is now a finite union of $f^{-1}(O_i)'s$ (where the O_i's are drawn from our original open cover of $f(U)$). This tells us that $f(U) \subseteq f(\cup U_i)$. But each $U_i = f^{-1}(O_i)$ for some O_i in the original cover of $f(U)$. Since $f(f^{-1}(O_i)) \subseteq O_i$, we have $f(U) \subseteq \cup O_i$, where the union on the right is finite. I.e., we have a finite subcover of $\{O_i\}$.

Pf: (v) Suppose $f(S)$ is not connected. Then it is the union of two disjoint open sets $f(S) = O_1' \cup O_2'$. However, $f^{-1}(O_1' \cap O_2') = f^{-1}(O_1') \cap f^{-1}(O_2')$, and both are $\emptyset$ since O_1' and O_2' are disjoint. $s \subseteq f^{-1}(f(S))$ in general (with equality if f is injective, which we're not assuming). So $s \subseteq f^{-1}(O_1' \cup O_2') = f^{-1}(O_1') \cup f^{-1}(O_2')$. I.e., $s = (s \cap f^{-1}(O_1')) \cup (s \cap f^{-1}(O_2'))$. However, $f^{-1}(O_1')$ and $f^{-1}(O_2')$ are disjoint and open, so we have a decomposition into disjoint sets $(s \cap f^{-1}(O_1'))$ and $(s \cap f^{-1}(O_2'))$. Though not necessarily open in X, these are open in the subspace topology for S (which consists of all sets of the form $O \cap S$ with O open in the topology on X). So S is a disjoint union of two open sets in the subspace topology and isn't connected, contradicting our premise.

Since $\mathbb{R}$ (or any manifold, for that matter) is Hausdorff, any continuous map $f : K \to \mathbb{R}$ from a compact space K satisfies (iii).

Prop 2.2: Given topological space X and subspace Y, the following hold:

- (i) If X is Hausdorff, so is Y.
- (ii) If X is second-countable, so is Y.
- (iii) If X is compact and Y is closed in X, then Y is compact.
- (iv) If X is paracompact and Y is closed in X, then Y is paracompact.
- (v) If X is locally compact and Y is either open or closed in X, then Y is locally compact.

Pf: (i) Let $Y \subset X$, with X Hausdorff. Suppose $y_1, y_2 \in Y$. Since X is Hausdorff, there exist disjoint open sets $O_1, O_2 \subset Y$ containing y_1 and y_2. Then $Y \cap O_1$ and $Y \cap O_2$ are open in Y (by definition), disjoint, and contain y_1 and y_2. So Y is Hausdorff.

For (ii), see [3], theorem 30.2. For (iii), see [3], theorem 26.2. For (iv), see [3], theorem 41.2. For (v) see [3], corollary 29.3.

2.2. **Note on different topologies on the same set.**

Let's make a tiny detour to clarify an important point. The choice of topology makes a big difference in how we view things. Just as a choice of σ-algebra reflects our notion of which sets are measurable (regardless of the particular measure) and thus determines the measurable functions to and from other σ-algebras, our choice of topology reflects our notion of which sets are open and thus determines the continuous functions to and from other topological spaces. Two different topologies on the same set yield different notions of continuity.

For example, consider the set $X \equiv [0,1] \cap Q$ (with Q the rational numbers). If we endow X with the discrete topology $T = 2^X$, then every set is open, and the topology "sees" down to the level of individual points. In this topology, every point is isolated because it has an open neighborhood (itself) which contains no other points. On the other hand, consider the same set X but with the subspace topology it inherits from $\mathbb{R}$. In this case, there are no "smallest" sets — reflecting the absence of a partition basis (to be defined shortly) — and we "see" things in terms of open balls

.

We can endow X with a variety of metrics and pseudometrics, including the discrete metric (i.e. $d(x,x) = 0$ and $d(x,y) = 1$ for $x \neq y$) and the standard metric inherited from $\mathbb{R}$ (i.e. $d(x,y) = \sqrt{(x-y)^2}$).

X also admits the indiscrete pseudometric (i.e. $d(x,y) = 0$ for all x, y). We still have a notion of an "open ball" with the indiscrete pseudometric, but it is trivial. However, we'll focus on metrics rather than pseudometrics here (besides which, the indiscrete pseudometric is uninteresting).

Note that the term "metric" is commonly used to refer to two very different things: a distance function on a set or a Riemannian metric on a manifold (i.e. a specific type of $(0,2)$-tensor field on the manifold). We'll usually mean a distance function, but we'll be explicit when confusion is possible.

Any metric on a set X endows it with a corresponding topology called the **ball topology** or **metric topology**. A **ball** (aka "open ball") is a set $B_r(x) \equiv \{y \in X; d(x,y) < r\}$ for real $r > 0$ and $x \in X$. The collection of all balls plus the empty set (which we can consider the ball of radius 0 if we allow $r = 0$) is the basis for a topology (i.e. it generates the topology via arbitrary unions of balls). Note that different metrics can induce the same ball topology on X.

For example, any norm $||$ on a vector space induces a metric via $d(v,w) \equiv |v - w|$. However, all norms on a finite-dimensional vector space induce the same topology (see [4], proposition 2.1.10). I.e., the norms may induce distinct metrics but all the resulting ball topologies are homeomorphic.

There is ambiguity in how the term "metric space" is used. Some people use it to mean just a set with a metric, while others include the induced topology as well. "Space" usually refers to either a vector space or a topological space, so we'll prefer "metric set" for a set with just the metric and "metric space" for that same object as a topological space with the induced ball topology.

A topological space is **metrizable** if there exists a metric space to which it is homeomorphic. We'll say that a topology T on X is **compatible with a metric** d on X if (T, X) is homeomorphic to (T_d, X), where T_d denotes the ball topology of d. For all practical purposes, this means that T and T_d are the same.

In our example above, we would say that the topology induced by any norm on a finite-dimensional vector space is compatible with that of every other norm on that space.

Note that any subset of a metric set inherits the metric in the obvious way.

Prop 2.3: Let d be a metric on X, let $Y \subset X$, and let T be the ball topology of d on X. Then the subspace topology on Y (inherited from T) is the same as the ball topology on Y obtained from $d|_Y$.

I.e., it doesn't matter whether we restrict d and then take the ball topology or take the ball topology and then go to the subspace topology.

Pf: Denote by T' the subspace topology on Y and by T'' the ball-topology on Y obtained from $d|_Y$. Consider the ball $B'_r(y)$ in Y, where the $'$ denotes that we are taking it based on $d|_Y$. It consists of all points $y' \in Y$ s.t. $d(y', y) < r$. Since X contains more points, $B'_r(y) \subseteq B_r(y)$, where the latter is the corresponding ball in X. In fact, $B'_r(y) = B_r(y) \cap Y$. Any open set $O' \in T'$ is of the form $O' = O \cap Y$ for some $O \in T$. However, any $O \in T$ can be written as a union of open balls (which form a basis for T): $O = \cup B_i$. Since $Y \cap O = Y \cap (\cup B_i) = \cup (Y \cap B_i) = \cup B'_i \in T''$, we have $T' \subseteq T''$. Consider any $O'' \in T''$. It is a union of basis elements $O'' = \cup B'_i = \cup (Y \cap B_i) = Y \cap (\cup B_i) \in T'$. I.e., $T' = T''$.

Returning to $X = Q \cap [0,1]$, consider both the discrete and inherited Euclidean metrics we mentioned. Under the discrete metric, every point is an open ball, and the topology has as basis X itself. I.e., the discrete metric induces the discrete topology. Any function from a discrete topology to another topology is continuous.

Since every set in the discrete topology is open, the inverse image of any open set in Y is open. Basically, the discrete topology trivializes the notion of continuity. As mentioned, it views things in terms of points.

On the other hand, the inherited Euclidean metric on X induces nothing other than the subspace topology (relative to the standard topology on $\mathbb{R}$). This isn't surprising, since the standard topology on $\mathbb{R}$ is just the Euclidean metric topology. The notion of open ball on X in the subspace topology is just $(a,b) \cap [0,1] \cap Q$ for some open ball $(a,b) \subseteq \mathbb{R}$. The subspace topology sees things in terms of open intervals of rational

numbers. Since Q is dense in $\mathbb{R}$ (and thus $Q \cap [0,1]$ is dense in $[0,1]$), we can view this in another light: the subspace topology sees $[0,1] \cap Q$ as approximating the standard topology on $[0,1]$. In the subspace topology, X has no isolated points. As far as it is concerned, standard open balls are what we traffic in, and the notion of continuity to and from this topological space reflects that.

The discrete topology is generated by the discrete metric and the inherited ball topology is generated by the inherited balls (i.e. $X \cap B_r(x)$), which also are the balls of the inherited metric. I.e., the subspace topology is compatible with the inherited standard metric. However, the subspace topology is *not* compatible with the discrete metric and the discrete topology is *not* compatible with the inherited metric.

We mentioned that (in some vague sense) a topology "sees" down to the level of open sets. Under certain circumstances, this can be made concrete, and we can think of a topology as giving us (again, from the standpoint of continuity), a precise notion of "resolution". We've already seen that this is true in the case of the discrete topology, where the resolution is the individual elements of X. This can be broadened somewhat.

Let T be a topology on X. A subset $B \subseteq T$ is a **basis for** T if every element $O \in T$ is a union of elements of B. A **partition basis** is a disjoint basis.

Suppose B is any disjoint basis for T. Since $X \in T$, X itself must be a union of elements of B. These elements are disjoint and thus form a partition of X. Hence the name.

If T has a partition basis, then, from the standpoint of continuity, there is no information below the level of the basis sets. Each open set has a unique expression as a disjoint union of these. As a partition of X, B induces an equivalence relation $\sim_B$, and it is easy to see that $\sim_B$ is compatible with T in the sense that each open set of T is a union of whole equivalence classes. In a way, we can view X as topologically "the same" as the quotient space $X' \equiv X/\sim_B$ under the corresponding quotient topology $T' \equiv T/\sim_B$ — but not in the sense of homeomorphism. T' is just the discrete topology on X' (i.e. $T' = 2^{X'}$). This does *not* make (X,T) homeomorphic to the discrete topology (X',T'), but T nonetheless behaves like a discrete topology. Another way to think of this is that X is a union of connected components, each an element of B, and has no open sets smaller than these.

The quotient topology takes a little care to define properly, and we won't do so here. See [3], section 22 for the details of its construction.

Note that if, instead of these clean-cut cases, we have something more complicated — say a set with two distinct structures — we very well could end up with two naturally-induced and possibly incompatible topologies. Usually, we would constrain the structures to be compatible to the extent of inducing the same topology. However, in general we could find ourselves forced to choose between distinct natural topologies. As discussed earlier, the physical assumptions would either have to constrain the structures to be compatible or give us a reason to favor one induced topology over the other(s) for our purpose.

2.3. **Some useful definitions and results for manifolds.**

We will be dealing with manifolds a lot, so let's review some relevant facts about them.

Recall that a FOO n-manifold M is a topological space that locally looks like K^n for either $K = \mathbb{R}$ or $K = \mathbb{C}$ and exhibits "FOO"-differentiability in a sense we will describe below. We say the manifold is of type FOO and is real or complex (according to K) and has dimension n.

FOO is the overlap condition that we'll describe shortly. If $K = \mathbb{R}$, FOO can be "topological" (aka continuous or C^0) or "k-differentiable" (aka k-times differentiable or C^k) or "smooth" (aka C^∞) or "real analytic" (aka C^ω). If $K = \mathbb{C}$, FOO can be "topological" or "complex analytic" (aka holomorphic, and also denoted C^ω) if $K = \mathbb{C}$.

We'll only consider finite-dimensional manifolds here, so n is finite.

Because FOO is specific to the choice of field, we'll treat the specification of K as part of the information in FOO.

If T is our topology on M, the "manifold structure" takes the form of an "atlas" of "charts", each of the form (U, ψ) where U is an open subset of M and ψ is a homeomorphism $\psi : U \to K^n$ s.t. (i) the charts cover M and (ii) on any overlap of two charts (U, ψ) and (U', ψ'), the overlap map $\phi = \psi' \circ \psi^{-1} : K^n \to K^n$ is suitably differentiable (in the usual sense) and has suitably differentiable inverse.

ψ is often defined going the other way instead (i.e. $K^n \to U$). Since ψ is a homeomorphism, this only affects the notation.

It is easy to see that ϕ is a homeomorphism on K^n.

"suitably differentiable" is a term we will use a lot for convenience. It just means that the condition FOO is obeyed in K^n by the overlap map. If FOO is "topological", then the condition is automatically satisfied since we have continuity in each direction. If FOO is "k-differentiable", we require that every overlap map be k-times differentiable in the usual sense, and so on.

Charts are also called "coordinate systems" or "choices of coordinates" or just "local coordinates".

We'll often drop the "FOO" from now on, with the implicit understanding that a dependence on the relevant FOO is attached to all manifold-related definitions. The notions of atlas, maximal atlas, chart, etc all depend on FOO.

A **differential structure** on M is a maximal atlas, obtained from any atlas by including all compatible charts. It is easy to show that there is a unique maximal atlas containing any given atlas.

Again, the notion of "differential structure" and "maximal atlas" depend on the choice of FOO.

The overlap condition is not the only constraint on the charts in a differential structure (or atlas or maximal atlas). A chart requires a homeomorphism ψ, so this too is a constraint. If a particular open set is not homeomorphic to K^n, then it can appear in no chart. For example, there is no chart containing all of S^n. A given open set U may appear in no charts or many charts. If it appears in one chart (U, ψ), it appears in many, because any suitably differentiable function $f : \psi(U) \to \psi(U)$ with suitably differentiable inverse yields a new chart (U, ψ') via $\psi' \equiv f \circ \psi$.

Note that a topological manifold is not just a topological space. It carries a maximal atlas, consisting of all open sets homeomorphic to K^n and all such homeomorphisms of those sets.

We can treat any FOO n-manifold as a FOO' n-manifold if FOO' is a weaker condition for the same K.

Technically, the maximal atlas expands when we weaken the condition. However, the new maximal atlas clearly contains the old one — so any existing charts remain in play.

I.e. every manifold is a topological manifold, every k-differentiable manifold is a $(k-1)$-differentiable manifold, every smooth manifold is a k-differentiable manifold (for any finite k), and every real-analytic manifold is a smooth manifold.

Note that there exists only one topological maximal atlas, and every FOO-atlas is a subset of it. The topological overlap condition is automatically satisfied by any two charts (i.e. there *is* no topological overlap constraint). All that the topological atlas does is collect all possible local homeomorphisms to K^n. Only when we start imposing differentiability overlap conditions, do some charts become incompatible with one another.

Prop 2.4: Given any real k-differentiable n-manifold, the differential structure contains a unique smooth differential structure.

I.e., if we're differentiable at all, we can always work in charts with the smooth overlap condition. This means that we lose no generality by considering only topological, smooth, and analytic real manifolds. It is the reason that many treatments of differential geometry focus on smooth manifolds.

See [5], theorem 2.9 of section 2.2 (page 51).

Suppose we have a FOO n-manifold M (with topology T) and a FOO n'-manifold M' (with topology T'). A FOO map between M and M' is a function $f : M \to M'$ between the underlying sets which is continuous and s.t. for each chart (U, ψ) of M and each chart (U', ψ') of M', the map $\psi' \circ f \circ \psi^{-1} : K^n \to K^{n'}$ is of type FOO.

For a topological manifold, FOO is continuity and is automatically satisfied by any f that is continuous relative to T and T'.

A "diffeomorphism" between FOO n-manifold M and FOO n-manifold M' is an invertible FOO map f whose inverse is a FOO map. From the standpoint of differential geometry, diffeomorphic manifolds are the same.

For topological manifolds, a diffeomorphism is just a homeomorphism between T and T'.

A given topological space may admit no differential structure of type FOO or a single such differential structure or multiple non-diffeomorphic structures.

$\mathbb{R}^2 \cup \mathbb{R}$ (i.e. the $x - y$ plane plus the z-axis) admits none since part of it is locally homeomorphic to $\mathbb{R}^2$ and part of it is locally homeomorphic to $\mathbb{R}$ (and around the origin it is not locally homeomorphic to any $\mathbb{R}^n$). As an example at the other extreme, $\mathbb{R}^4$ admits infinitely many non-diffeomorphic smooth structures.

A real differentiable manifold M (i.e. a manifold with any real FOO except "topological") carries a host of automatically-generated companion machinery. These include the following:

- a de Rham complex (i.e. a graded algebra of tensor fields, a graded algebra of differential forms, a wedge product, and an exterior derivative).
- a tangent vector bundle TM.
- a cotangent vector bundle T^*M, along with a tautological 1-form and tautological 2-form on T^*M.
- tensor vector bundles of all (p, q).
- differential-form bundles of all p.
- a frame bundle LM (which serves as the principal bundle for all the bundles thus described, under the obvious action of $GL(K^n)$), along with a tautological solder form on LM.
- a (very large) Lie Algebra of vector fields, with $[v, w](f) = v(w(f)) - w(v(f))$ as the relevant Lie bracket.
- for each vector field, a Lie derivative that maps tensor fields to tensor fields.
- a group $Diff(M)$ of diffeomorphisms $M \to M$.
- for each vector field, an interior product i_v that maps forms to forms.
- a notion of orientability (or not) and, if orientable, a set of 2^m orientations (where m is the number of connected components).
- if orientable, a notion of volume form and a nonempty set of such forms.

Note that all the fiber bundles listed, other than LM, are vector bundles rather than plain fiber bundles.

Recall that a vector bundle doesn't just have a vector space as fiber. For a vector bundle, (i) the overlap condition of the fiber bundle atlas also requires linearity, (ii) the notion of morphism involves not just a fiber-preserving suitably-differentiable map, but also one linear on each fiber, and (iii) the notion of isomorphism involves not just a fiber-preserving diffeomorphism but one that is a vector-space isomorphism on each fiber.

A topological manifold alone induces none of the above machinery.

We require differentiability to define tangent vectors, and all else follows from this.

Strictly speaking, orientability and orientation can be defined for topological manifolds as well as differentiable manifolds, but not nearly as easily. Instead of Jacobians of the overlap maps, we must work with homology groups of the manifold minus each point. We won't go into it here (see [6], p. 233 of section 3.3 for details). Bear in mind that in either case, orientability is a property of the manifold, but orientation is a choice for each of its connected components. An orientable manifold with m connected components has 2^m possible orientations.

Almost everybody defines manifolds to be at least Hausdorff and paracompact, and we'll follow this convention. Every manifold is locally compact, even if we do not take it to be paracompact. This follows because it locally looks like $\mathbb{R}^n$ or $\mathbb{C}^n$, both of which are locally compact.

Some people replace paracompact with second-countable in the definition. This is a stricter requirement, but not in any way that we'll care about. Proposition 2.5 below tells us that a (Hausdorff and paracompact) manifold is second-countable iff it has a countable number of connected components. In particular, any connected manifold is second-countable, as is any compact manifold. We'll eventually want second-countability, but we'll impose this as a requirement rather than as a part of the definition.

Paracompactness guarantees that the manifold admits a partition of unity, necessary for defining integration. It almost never is used except in the context of Hausdorff spaces.

A manifold has any "locally-foo" property that Euclidean space has, because it locally looks like Euclidean space. Although they may seem local, neither Hausdorff nor paracompact are local properties. Every manifold is locally Hausdorff, but it is possible to omit Hausdorff from the definition and still have a meaningful notion of manifold.

Prop 2.5: A (Hausdorff) manifold is second-countable iff it is paracompact and has a countable number of connected components.

See [7], exercise 1-5 (or [8] for a detailed proof).

Warning: Intuitively, it may seem that any second-countable space X must have a countable set of connected components. The reasoning goes as follows: if X is a disjoint union of open sets, and each of these is a union of basis elements, then we need at least one basis element for each component. So if the basis is countable, there can be only a countable number of connected components. However, this is not true. For example, $X = \mathbb{R} - \mathbb{Q}$ is a subspace of $\mathbb{R}$. Since $\mathbb{R}$ is second-countable, so is its subspace X. However, X has an uncountable number of components (in fact, every point in X is a connected component). The connected components still form a disjoint cover of X. However, they are not open so they are not in the topology. As such, they needn't be expressable as unions of basis elements — and our argument breaks down. To ensure that the connected components are open, we need additional constraints. A manifold is locally-connected (because it is locally Euclidean) and Hausdorff, and these turn out to be enough to guarantee that the connected components are open. In that case, they must be unions of basis elements, and our argument for their countability works. In our counterexample, $\mathbb{R}$ is a manifold, but $\mathbb{R} - \mathbb{Q}$ is not — and this is the catch. As a subspace of $\mathbb{R}$, X is both Hausdorff and second-countable. However, unlike $\mathbb{R}$, X is not locally-connected.

We will find the following result about submanifolds useful:

Prop 2.6: The (nonempty) open sets of an n-manifold are its n-dimensional (embedded) submanifolds. If the manifold is orientable, all its open subsets are too.

See [7], proposition 5.1.

Unless otherwise stated, all manifolds we'll consider in these notes are smooth real manifolds.

2.3.1. *Integration.*

In essence, there is only one type of integral we know how to take in differential geometry: the integral of an n-form over the entirety of the n-manifold M. Why "in essence"? There are two important caveats here. The first involves orientability and orientation, which we'll have more to say about shortly. The second caveat is something glossed over in many treatments, but will be crucial for the Riesz representation

theorem. It involves our ability to take integrals of the form $\int_M f\nu$, where ν is an n-form and f is a *continuous* (rather than smooth) real function on M. These caveats aside, the key point is that we can only integrate (i) over the entire manifold and (ii) something involving a top-level form.

If we wish to take the integral of a $k < n$ form, we must first confine ourselves to a k-dimensional submanifold. This constitutes additional information, and tells us which points to confine ourselves to as well as the relevant k degrees of freedom to retain in their tangent spaces. Similarly, if we wish to integrate over an n-dimensional "chunk" of the manifold (ex. an open set), we must first codify that chunk as an n-dimensional submanifold and then integrate over it.

We won't fully develop the theory of integration here, but let's go over the basics. We know how to take the Lebesgue integral of a measurable function $f : \mathbb{R}^n \to \mathbb{R}$. We'll see in proposition 4.7 that every continuous function is measurable (though the converse does not hold). We therefore can take the Lebesgue integral of any continuous f.

The terminology surrounding integration can be misleading. There is a set of functions for which the integral is well-defined, and a subset of these for which it is finite. "Integrable" means having an integral which is finite, not just well-defined. There may be functions whose integrals are well-defined and infinite, but these are not called "integrable". This can be counterintuitive and a potential source of confusion. As a simple example, $f(x) = 1$ is not "integrable" on $\mathbb{R}$. The integral exists but is infinite. "Absolute integrable" means that $|f|$ is integrable, which implies that f is integrable as well.

The idea of integration on a manifold is to patch together Lebesgue integrals chart-by-chart. Since a smooth n-manifold M is locally diffeomorphic to $\mathbb{R}^n$, we just need to wrangle our integrand into $\mathbb{R}^n$ and integrate it there. We then take a sum over these local integrals.

This sounds simple but is much easier said than done. The charts are not disjoint, so we need to avoid redundant contributions from overlapping charts. Moreover, we must ensure that whatever method we use is independent of the set of charts (as long as they form at atlas). This is accomplished via something called a partition of unity. We'll assume that the manifold is orientable and that we've picked an orientation.

As we'll mention below, it is possible to integrate something called a "density" rather than a form on a non-orientable manifold, but this is beyond the scope of our present discussion.

If we are integrating a volume-form (or something involving one, such as $\int_M f\nu$), we'll pick the orientation that makes ν positive.

Given an oriented atlas (defined below), we define a set of functions $\phi_i : M \to [0, 1]$, indexed by the charts of the atlas (i.e. one ϕ_i per chart) and s.t. (i) each ϕ_i has support only in the corresponding chart, (ii) the ϕ_i are smooth, and (iii) at every point x, $\sum \phi_i(x) = 1$. In chart (U_i, ψ_i), the n-form ν can be written in local coordinates as $g_i(x_1, \dots x_n)dx^1 \wedge \dots \wedge dx^n$ for some smooth g_i. We define $I_i \equiv \int (\phi \circ \psi^{-1})(x_1, \dots x_n) \cdot g_i(x_1, \dots x_n)dx_1 \cdots \dots dx_n$, where the expression on the right is the ordinary Lebesgue integral (over all of $\mathbb{R}^n$) of the function $(\phi \circ \psi^{-1}) \cdot g_i : \mathbb{R}^n \to \mathbb{R}$. Our global integral then is $\int_M \nu \equiv \sum_i I_i$, where the sum is over all the charts in the oriented atlas.

This is a vast oversimplification. There are lots of technical details involved, such as proving that the choice of the partition of unity doesn't matter and that such a partition of unity even exists. The definition of the sum requires some clarification as well, since the relevant atlas may be uncountable. Because a manifold is locally compact, we can choose our partition of unity in a way where the sum contains only finitely many nonzero terms in the neighborhood of any point. We won't go into the details here.

If we confine ourselves to the integration of n-forms, we run into a problem. As mentioned above, an n-form can be written $g_i(x_1, \dots x_n)dx^1 \wedge \dots \wedge dx^n$ in local coordinates — but with g_i a *smooth* function. Even though every smooth atlas is also a C^k atlas (for every $k > 0$), the best we can do is loosen the requirement to differentiability of g_i. There is no notion of a differential form on a topological manifold, so we can't

take it further than this. Intuitively, that doesn't feel right. We can Lebesgue integrate non-differentiable functions on $\mathbb{R}^n$. We can even integrate discontinuous ones, as long as they are measurable. Nor is this purely a conceptual complaint. We'll need the ability to integrate continuous functions when discussing the Riesz representation theorem.

Technically, for the Riesz theorem we'll only need to be able to integrate continuous functions with compact support, but compactness of the support isn't the obstacle here. We need to extend our notion of integration to continuous real functions on M.

Though it's a point that tends to be glossed over in many treatments, the partition-of-unity approach works extends just fine to the integration of continuous functions. For $\int_M f\nu$, we simply use $I_i \equiv \int(\phi \circ \psi_i^{-1}) \cdot g_i \cdot (f \circ \psi_i^{-1}) dx_1 \cdots \ldots dx_n$.

Of course, just saying that this works doesn't mean it works. However, it is possible to show that this integral is well-defined at the manifold level (i.e. independent of the choice of charts and the partition of unity). The non-smoothness of f doesn't pose an obstruction. See [9], section 25 for a detailed development, albeit using a slightly different approach.

By a similar token, this same method should extend to any f for which $(f \circ \psi_i^{-1})$ is measurable in every chart. In principle, this could include some discontinuous functions as well — though I've never seen it extended this way.

If we take $f = 1$, we get the ordinary n-form integral $\int_M \nu$. In the case where ν is a volume-form, this defines the "volume" of M imparted by ν.

2.3.2. *Volume forms and orientation.*

Suppose M is a smooth n-manifold with m connected components (where m may be infinite, even uncountably so). An **oriented atlas** of M is a subatlas of its maximal atlas s.t. the Jacobian of the overlap maps is positive everywhere. A **maximal oriented atlas** is an oriented atlas that contains all charts which satisfy this Jacobian condition. Obviously, all maximal oriented atlases are subatlases of the maximal atlas which comprises M's differential structure. M is called **orientable** iff it has an oriented atlas. If it is orientable, it has 2^m maximal oriented atlases. Each such maximal oriented atlas is called an **orientation** of M.

We can make this more precise for the case of infinite m. Let $CO(M)$ denote the set of connected components of M. Given a reference orientation $o \in CO(M)$, each other orientation is related by a map $CO(M) \to \{-1, +1\}$, where the value for the i^{th} component is the sign of the Jacobian of its charts relative to those of o. For finite m, we can think of this as an m-tuple of ± 1 (obviously, these are all $+1$ for o itself). Note that we cannot say that a component takes $+1$ or -1 in any absolute sense. The set of orientations and set of maps $CO(M) \to \{+1, -1\}$ are bijective, but there is no natural choice of bijection without additional information. If we pick a reference o, we have a specific bijection. As we will see, a choice of volume form fixes a specific o.

A volume form on a smooth orientable n-manifold is a positive (and thus nonvanishing) n-form. A manifold has a volume form iff it is orientable, in which case it has infinitely many.

Some treatments define a volume form as nondegenerate. The concepts of "nondegeneracy" and "nonvanishing" differ for general k-forms (with $k < n$), but not for n-forms.

What do we mean by positive? Consider a specific orientation o for the manifold. In a chart of o with local coordinates $(x_1 \dots x_n)$, any n-form can be written $f(x_1, \dots, x_n)dx_1 \wedge \cdots \wedge dx_n$ for some smooth f. Nonvanishing on the chart means that f never vanishes, and positive on the chart means that $f > 0$ everywhere. A nonvanishing form is nonvanishing in every chart of the maximal atlas and a positive form is positive in every chart of the chosen o. Note that a nonvanishing form on a chart must have the same sign everywhere; otherwise, by continuity, it would have to pass through 0, which means it would vanish somewhere. Along with the Jacobian condition, this means that a nonvanishing form must have the same sign on all charts of an oriented atlas. I.e., in a given orientation, the values of a nonvanishing form are all positive or all negative on each component. If there is more than one component, "positive" means positive on all of them.

There is some ambiguity in the term "volume form" as stated. An n-form can only be positive in a single orientation. Do we say that ν is a "volume form" only in that orientation or do we say that an n-form ν is a "volume form" if there is any orientation which makes it positive? The answer is: both. We say that an n-form is a volume form iff there exists an orientation that makes it positive — and we always assume we are working in that orientation.

Our integral as defined above depends on our choice of orientation. It is not hard to see that if M has m connected components M_i, then $\int_M f\nu = \sum_i \int_{M_i} f|_{M_i}\nu|_{M_i}$ in our chosen orientation o. I.e., we can perform our integration procedure on each component submanifold and sum the resulting integrals. It also is easy to see that if we vary our orientation we get up to 2^m distinct values for the integral this way. Using o as our reference orientation, if we switch to orientation o' there is a corresponding $\alpha_{o,o'} : CO(M) \to \{-1, +1\}$ that tells us the sign of each component under o' relative to o. In orientation o', our integral is of the form $\sum_i \alpha_{o,o'}(M_i) \int_{M_i} f|_{M_i}\nu|_{M_i}$.

If we are taking $\int_M \nu$ or $\int_M f\nu$ for a volume form ν, then we have a specific preferred orientation — and our meaning is clear. However, it is often desirable to avoid a dependence on the orientation. There are two common ways to do so: pseudo-forms and densities. Densities actually allow us to perform integration on non-orientable manifolds as well, and they crop up a lot (for other reasons) in General Relativity. We won't go into them here. A pseudo-form is a map from orientations to volume forms that allows us to pick a materially equivalent volume form for each orientation.

It is not hard to see that an n-form ν on M can be written as a sum $\nu = \sum_{M_i} \nu_i$, where $\nu_i = \nu|_{M_i}$ on M_i and is zero elsewhere. I.e., we simply break ν into its parts on the components. If ν is a volume form with associated orientation o, and $\alpha_{o,o'} : CO(M) \to \{-1, +1\}$ is the sign map for orientation o', then $\sum \alpha_{o,o'}(M_i)\nu_i$ is a volume form with orientation o'. Using this, we obtain a map from $CO(M)$ to volume forms which are materially identical on each component but differ in the orientation relative to which they are positive. We then can define integration in an orientation-independent manner by picking out the relevant volume form for whatever orientation we choose. The integral will be the same in every case. It is easy to see that both $\int_M \nu$ and $\int_M f\nu$ can be defined in a manner which gives the same result as the ordinary method we described above.

Prop 2.7: Let ν be a volume form on n-manifold M. For any open $U \subset M$, $\nu|_U$ is a volume form on U.

Pf: As mentioned in proposition 2.6, the open subsets of an n-manifold are precisely its embedded n-submanifolds. Let $i : U \to M$ be the embedding map. Since U is an embedded n-submanifold, the push-forward $i_* : T_qU \to T_qU$ at each point has full rank and we don't need to choose which directions to retain in the tangent space. I.e., we can just restrict ν to U without any complications. As an n-manifold, U automatically inherits orientability and orientation from M. Since positivity of the n-form is pointwise, this is preserved under restriction. Formally, if $i : U \to M$ is the inclusion map, the corresponding n-form on U is the pullback $i^*\nu$.

Restriction of this sort only works for full-dimension embedded submanifolds. If we restrict to a k-dimensional submanifold N of n-manifold M (with $k < n$), the tangent space is of lower dimension. We could reduce a volume form ν on M to one on N via contraction (aka the interior product) $i_V\nu$ for some vector field V on M. However, this depends on the choice of V. We thus would need additional information that provides a natural choice of V. Even worse, we would not be guaranteed to inherit orientability, let alone an orientation. For an n-dimensional manifold, we're simply taking an open subset of the points on the manifold. Each point has the same tangent space as in the original manifold, and these issues do not arise.

If (smooth n-manifold) M is orientable, then so is every n-submanifold. Moreover, each orientation on M is directly inherited by each n-submanifold. As a result, the orientability and all the orientations of M are trivially inherited by every open set of M.

For any continuous function f and any volume form ν on M, it follows that we can compute $\int_U f\nu$ on any open subset $U \subset M$, not just on M itself. Denoting the inclusion map $i : U \to M$, the integral is $\int_U f|_U \cdot i^*\nu$. In particular, we may compute the volume $\int_U i^*\nu$ of U that is imposed by ν.

3. The sets of microstates: Ω and Ω_E.

Let Ω be the complete set of microstates for our system, and let $H : \Omega \to \mathbb{R}$ be our Hamiltonian, taking each microstate to its energy. The set of all energies $S_H \equiv H(\Omega)$ is called the **energy spectrum**, though we'll just refer to it as the **spectrum**. We'll say that the spectrum is "discrete" if it consists solely of topologically isolated points in $\mathbb{R}$ (i.e. it is a union of discrete points), and we'll say that it is "continuous" if it consists of a union of intervals in $\mathbb{R}$. It is also possible for the spectrum to have both discrete and continuous regions.

By "interval", we'll mean open, closed, or half-open/half-closed, and possibly infinite at one or both (open) ends — with the exclusion of a single point closed interval, which is a discrete point. All Borel sets of $\mathbb{R}$ are unions of intervals. Although it may be tempting to imagine that a union of intervals is equivalent to a subset of $\mathbb{R}$ with no topologically isolated points, it is not. $\mathbb{Q}$ is an example of a set that has no topologically isolated points but cannot be written as a union of intervals.

Most common classical systems (point-particle systems, rigid-bodies, etc) have continuous spectra, though discrete-like behavior can nonetheless arise asymptotically. The most common examples of actual discrete spectra arise in quantum mechanics. The quintessential case is the quantum harmonic oscillator, whose spectrum is $\{\hbar\omega(n + \frac{1}{2}); n \in \mathbb{Z}\}$. More generally, consider the quantum mechanical spectrum of a single particle in an external potential in $\mathbb{R}^n$. The spectrum is discrete iff $V \to +\infty$ as $|x| \to \infty$ (see [10]). There is an analogous, but more complicated, condition known as the Molchanov criterion for a particle semi-bounded below. In either case, if the relevant condition for V is not met, it is possible to have a mixed spectrum.

The terms "discrete" and "continuous" spectrum arise from the visual appearance of the graph in spectrometry, rather than the mathematical nature of the Hamiltonian in those regions. The Hamiltonian is always assumed to be continuous.

The AEAP requires us to impose a uniform probability distribution on the relevant set of states. In the microcanonical ensemble, the relevant set consists of all microstates with the fixed energy in question. I.e. $H^{-1}(E)$ for some E under consideration. For continuous spectra, we cannot measure a specific E, so we typically speak of an interval of E's. We are free to choose the type of interval as we please, as long as it is small enough to represent the notion of "fixed" E for all practical purposes. The most convenient choice is an open interval $\delta_E \equiv (E - \epsilon, E + \epsilon)$ for some suitably small ϵ (and allowing the choice of "suitably small ϵ" to depend on E).

This works for discrete spectra too, since $H^{-1}(\delta_E) = H^{-1}(\{E\})$ if we choose ϵ to be suitably small. We'll therefore employ it without reservation for both the discrete and continuous cases, though we'll sometimes also just say $H^{-1}(E)$ when discrete.

Bear in mind that "discrete" and "continuous" refer to the subsets of $\mathbb{R}$ which are in S_H, *not* to the sets of states which map to them.

We'll denote by $\Omega_E \equiv H^{-1}(\delta_E)$ (for whatever ϵ is suitable to that E) the relevant set of states for which the AEAP demands uniform probabilities.

3.1. Ω and Ω_E as topological spaces.

In classical mechanics, the state space is the "phase space" associated with a configuration space. We start with a **configuration manifold** Q, typically defined in terms of some set of generalized coordinates

$\{q_1, \ldots, q_n\}$ subject to various constraints.

I.e. it is a hypersurface in some $\mathbb{R}^n$.

Note that there are multiple physical assumptions involved here: (i) that Q is a manifold and (ii+) how to construct Q for a given physical system. We won't delve into the latter because it is not relevant to our present machinery. Once (i) is satisfied, the AEAP (and statistical mechanics) is agnostic to the details of how and why a particular Q was chosen as the relevant configuration manifold.

We're "agnostic" to (ii+) in the abstract sense only. The details of the remaining assumptions (ex. symmetries) can be very useful in understanding a given system and calculating with it.

The **phase space** associated with a configuration manifold Q is its cotangent bundle T^*Q. A core assumption of classical mechanics is that this is the complete set of microstates of our classical system (i.e., $\Omega = T^*Q$). Q (and thus T^*Q) is actually taken to be a smooth, rather than just topological, manifold, and T^*Q has a lot of nice properties and structure of its own.

Technically, we only require Q to be a C^2 or C^3 manifold in most of classical mechanics. The embedding theorem 2.4 tells us that we may assume it is smooth without any loss of generality.

Since the overall set of microstates Ω is a topological space, we can meaningfully speak of whether the Hamiltonian H is continuous. We will take the continuity of H as a physical assumption.

Classical mechanics actually requires H to be at least C^1 (i.e. once-differentiable).

The topology on Ω endows each Ω_E with the subspace topology.

Recall that for topological space (Ω, T), the subspace topology on $\Omega_E \subset \Omega$ is just $\{O_i \cap \Omega_E; O_i \in T\}$. I.e., we just restrict each open set to Ω_E (possibly reducing it to the empty set).

Since (i) we've taken δ_E to be an open interval in $\mathbb{R}$ and (ii) H is continuous by physical assumption, $\Omega_E = H^{-1}(\delta_E)$ is open in Ω.

Note that it is not enough that we *have* a topological space. We need it to be physically meaningful. Topology embodies our notion of continuity, so the relevant notion of continuity for H is that associated with the topology on T^*Q, as inherited from Q. In the classical case, we are handed this topology as part of the definition of the manifold Q.

We've now established (or just stated) that:

(i) Classical mechanics hands us (as a physical assumption) a manifold Q, termed the configuration space, for any given system.
(ii) Classical mechanics tells us (as a physical assumption) that the overall set of microstates Ω for that system is the cotangent bundle T^*Q, which has a topology and manifold stucture derived from that of Q.
(iii) Classical mechanics hands us (as a physical assumption) a Hamiltonian $H : \Omega \to \mathbb{R}$ for the system, continuous relative to this induced topology on Ω (and the standard topology on $\mathbb{R}$).
(iv) As a consequence, the set of microstates $\Omega_E \subset \Omega$ relevant to the AEAP is a topological space with the subspace topology.

3.2. **Properties of Ω_E.**

As a manifold, the overall state space Ω is Hausdorff, paracompact, and locally compact. However, it is never compact.

Ω is a cotangent bundle, and it is easy to see that a vector bundle cannot be compact.

We saw that Ω_E is open in Ω since δ_E is open in $\mathbb{R}$ and H is continuous. Proposition 2.6 tells us that Ω_E is an embedded submanifold of Ω of the same dimension as Ω.

Every cotangent bundle is orientable, so both Ω and Ω_E are orientable and thus admit volume forms.

As will be discussed later, a cotangent bundle is a symplectic manifold and every symplectic manifold is orientable. Note that this does not require Q itself to be orientable.

Prop 3.1: If there is an isolated energy E in the spectrum, then Ω_E is a union of connected components of Ω (i.e., it does not contain just a piece of some component).

Pf: An isolated energy E has the property that $(E-\epsilon', E+\epsilon') \cap S_H = \{E\}$ for small enough ϵ'. Pick this as the ϵ for our δ_E. Pick any nonempty open set $O \subset S_H$ that does not contain E. Since $\delta_E \cap S_H = \{E\}$, $O \cap \delta_E = \emptyset$. So $H^{-1}(O) \cap H^{-1}(E) = \emptyset$. Now, consider the union of all such $H^{-1}(O)$'s: $U \equiv \cup_{O; E \notin O} H^{-1}(O)$. Each $H^{-1}(O)$ is open in Ω, so their union U is open too. Clearly, U and $H^{-1}(E)$ form a two-set disjoint open cover of Ω. By the definition of connectivity, this means that each must be a union of connected components.

This tells us that if there is a discrete part of the spectrum, the corresponding set of states must be a union of connected components of Ω.

There must be at least one such connected component for each discrete E, but there can be more than one. Each connected component is both open and closed in the topology on Ω. Since Ω is a $2n$-manifold (with n the dimension of Q), each component is a $2n$-dimensional embedded submanifold.

4. Review of some relevant measure theory

4.1. **Basics.** Recall that a **σ-algebra** on S is a set Σ of subsets of S that contains S and is closed under complements and countable unions. A **σ-subalgebra** of Σ is a subset of Σ that is a σ-algebra in its own right.

Closure under complements and countable unions implies that Σ contains $\emptyset$ and is closed under countable intersections as well. In fact, we can swap countable unions with countable intersections in its definition.

The elements of S are called **elementary events** and the elements of Σ are called **events** (or **measurable sets**).

We'll say that a σ-algebra Σ has $Z \subseteq \Sigma$ as a **subbasis** (or that it is **generated** by Z) if it is the smallest σ-algebra containing Z. Equivalently, every element of Σ can be constructed from complements and countable unions of elements of Z.

We'll say that a σ-algebra Σ has $B \subseteq \Sigma$ as a **basis** if every element of Σ can be expressed as a countable union of elements of B. Obviously, B forms a cover of S, since $S \in \Sigma$. If the elements of B are pairwise-disjoint (and thus form a partition of S), we'll say that B is a **partition basis**.

It is easy to show that (i) any partition basis is countable, (ii) if a partition basis exists, it is unique, and (iii) a σ-algebra can have any cardinality other than $\aleph_0$ (i.e. there is no such thing as a denumerably-infinite σ-algebra). Note that Σ can have uncountable bases and can contain uncountable partitions. It just cannot have an uncountable partition basis. For example, the powerset $\Sigma = 2^S$ has the singletized version of S as a partition but not a basis. Any σ-algebra has itself as a basis, but not as a partition.

Recall that in the topological case a disjoint basis represents a finest "resolution" below which there is no information from a topological standpoint. A partitition basis for a σ-algebra likewise represents a finest resolution below which there is no information from a probability standpoint. We lose no information if we replace (S, Σ) with $(B, 2^B)$, treating B as the effective set of elementary events and using its power-set σ-algebra.

It is easy to see that any finite Σ has a partition basis (consisting of all intersections of its elements). Since a σ-algebra can't have cardinality $\aleph_0$, if Σ has an infinite partition basis, $|\Sigma| = \beth_1$.

A measure on Σ is a function $\mu : \Sigma \to \mathbb{R}$ that is nonnegative and countably additive, meaning that for any countable set of disjoint events $\{s_i\}$, $\mu(\cup s_i) = \sum \mu(s_i)$. A **finite measure** has $\mu(S) < \infty$ (which implies that $\mu(s) < \infty$ for all $s \in \Sigma$ as well).

Nonnegativity and countable additivity imply a large number of other properties of μ as well, including $\mu(\emptyset) = 0$, monotonicity ($s \subset s' \implies \mu(s) \leq \mu(s')$), and subadditivity ($\mu(s \cup s') \leq \mu(s) + \mu(s')$). Note that it is quite possible to have $\mu(s) = 0$ for $s \neq \emptyset$. For a detailed discussion of various properties of measures, as well as which subsets of these properties can serve as alternate definitions, see [11], chapter 7.

It is easy to see that if Σ has a partition basis, then a measure is uniquely determined by its values on that basis.

A **probability measure** has $\mu(S) = 1$. A **probability field** is a triplet (S, Σ, μ), where μ is a probability measure on Σ.

We will use the terms "probability field" and "probability measure" interchangeably, with the choice of S and Σ implicit.

Prop 4.1: Given (S, Σ, μ) and σ-subalgebra $\Sigma' \subset \Sigma$, (i) the measure μ restricts to a measure on Σ' and (ii) finite measures restrict to finite measures, infinite measures restrict to infinite measures, and probability measures restrict to probability measures.

Pf: (i) Nonnegativity and countable additivity aren't damaged when we restrict ourselves to Σ'. In fact, they are easier to satisfy (i.e., there are functions μ which fail to be measures on Σ but restrict to measures on Σ'). (ii) Since $S \in \Sigma'$, the finiteness and total volume of the measure are unchanged.

Given (S, Σ) and (S', Σ'), we say that $f : S \to S'$ is a **measurable function** if $f^{-1}(s') \in \Sigma$ for every $s' \in \Sigma'$.

This is analagous to the definition of continuity, but with open sets replaced by measurable sets.

Measurability is relative to the σ-algebras at both ends, but we'll often omit mention of them if the choice is clear. For example, suppose we are considering $f : S \to \mathbb{R}$, with the Borel algebra Σ_{Bor} on $\mathbb{R}$ and some σ-algebra Σ on S. We may say that f is "measurable relative to Σ" or just "measurable", with the implicit understanding that this means it is measurable relative to (S, Σ) and $(\mathbb{R}, \Sigma_{Bor})$.

Prop 4.2: Let f be measurable relative to (S, Σ) and (S', Σ'). If Σ is a σ-subalgebra of Σ_1 on S and Σ'_2 is a σ-subalgebra of Σ' on S', then f is measurable relative to (i) (S, Σ) and (S', Σ'_2) and (ii) (S, Σ_1) and (S', Σ').

I.e., if we expand the source Σ or shrink the destination Σ', measurability is preserved.

Pf: (i) Let $s' \in \Sigma'_2$. Since $\Sigma'_2 \subset \Sigma'$, $s' \in \Sigma'$, and we know from f's measurability against Σ and Σ' that $f^{-1}(s') \in \Sigma$. (ii) Let $s' \in \Sigma'$. We know from f's measurability against Σ and Σ' that $f^{-1}(s') \in \Sigma$, so it also is in the larger σ-algebra Σ_1.

As the following proposition demonstrates, we need only test the measurability of a function on a subbasis at the target end.

Prop 4.3: Suppose we are given (S, Σ) and (S', Σ'), and let Σ' be generated by Z'. Then $f : S \to S'$ is measurable relative to Σ and Σ' iff $f^{-1}(s') \in \Sigma$ for all $s' \in Z'$.

I.e., it suffices to test Z' rather than all of Σ'.

Pf: Obviously, if f is measurable, then $f^{-1}(s') \in \Sigma$ for all $s' \in Z'$, since $s' \in Z'$ implies $s' \in \Sigma'$. Going the other way, suppose that $f^{-1}(s') \in \Sigma$ for all $s' \in Z'$. Since $f^{-1}(\cup s') = \cup f^{-1}(s')$, which is in Σ for countable unions, and $f^{-1}(\overline{s'}) = \overline{f^{-1}(s')} \in \Sigma$, and any element of Σ' is obtained from elements of Z' via such operations, $f^{-1}(s') \in \Sigma$ for all $s' \in \Sigma'$.

Note that $f^{-1}(Z') \subseteq \Sigma$, but it need not be a subbasis for Σ.

4.2. **Borel algebras.**

Every topology T has an associated **Borel algebra**, the smallest σ-algebra that contains T. In the Borel algebra of T, all open sets and all closed sets (and quite a few other sets) are measurable.

Prop 4.4: The Borel algebra on $\mathbb{R}$ (with the standard topology) is generated by any of the following sets: (i) all finite open intervals (x, y) with $x < y$, (ii) all finite closed intervals $[x, y]$ with $x < y$, (iii) all finite intervals of the form $(x, y]$ with $x < y$, (iv) all finite intervals of the form $[x, y)$ with $x < y$, (v) all half-open intervals of the form $(x, +\infty)$, (vi) all half-closed intervals of the form $[x, +\infty)$, (vii) all half-open intervals of the form $(-\infty, x)$, and (viii) all half-closed intervals of the form $(-\infty, x]$.

See [12], theorem 3.6 for (i)-(iv) and (viii). The rest follow readily from these. Ex. suppose (v) holds. Then we can obtain any $(x, y]$ via $(x, +\infty) \cap \overline{(y, +\infty)}$, and we then have (iii) which is a subbasis. Any superset (in Σ) of a subbasis is a subbasis too.

A measure on the Borel algebra (or larger) of T (with underlying set S) is **locally finite** if every $x \in S$ has an open neighborhood with finite measure.

Prop 4.5: If T is Hausdorff and μ is a locally finite measure on its Borel algebra, then every compact set has finite measure.

For a Hausdorff space, every compact set is closed. Since every closed set of T is in its Borel algebra, this includes every compact set. Therefore, referring to the measure of compact sets is valid. We often see Hausdorff as a requirement for a definition or theorem when there is a need to speak of the measure of compact sets. It ensures that compact sets are measurable.

Pf: Suppose μ is locally finite. Consider compact set K. Local finiteness tells us that each $x \in K$ has an open neighborhood O_x with finite measure. Together, these form an open cover $\{O_x; x \in K\}$ for K. Every open cover of a compact set has a finite subcover, so let $\{O_i\}$ denote such a finite subcover. $K \subseteq \cup O_i$, so (by monotonicity) $\mu(K) \leq \mu(\cup O_i)$ which (by subadditivity) is $\leq \sum \mu(O_i)$. However, each $\mu(O_i)$ is finite, so their sum is.

If T is not Hausdorff, then the argument above only applies to measurable compact sets. In that case, locally finite implies that every measurable compact set has finite measure. It follows that any locally finite measure on a compact space is finite, because the space itself is always contained in the topology and therefore in the Borel algebra.

Prop 4.6: If μ is a measure on the Borel algebra of a locally compact, Hausdorff space T such that every compact set has finite measure, then it is locally finite.

Pf: Consider $x \in S$. Since T is locally compact, x has an open neighborhood O_x that is contained in a compact set K_x. Therefore $\mu(O_x) \leq \mu(K_x)$, which is finite. So every point has an open neighborhood of finite measure.

I.e., for locally compact, Hausdorff spaces, "locally finite" and "every compact set has finite measure" are equivalent properties for a measure.

Warning: some treatments that confine themselves to Hausdorff, locally compact spaces incorrectly *define* locally finite via the compact set condition. In their case, the two are equivalent — but using that as the definition is misleading and should be avoided.

As the following proposition tells us, all continuous functions are measurable. However, not every measurable function is continuous.

Prop 4.7: Given topological spaces (X, T) and (X', T'), any continuous function $f : X \to X'$ is measurable relative to the Borel algebras of T and T'.

Pf: Let f be continuous and let Σ and Σ' be the Borel algebras of T and T'. For open $O' \in T'$, continuity tells us that $f^{-1}(O')$ is open in T and therefore measurable in Σ. Consider closed set C' under T'. The inverse image of a closed set under a continuous map is closed, so $f^{-1}(C')$ is closed under T and therefore is in Σ. Next, consider a countable union $S' \equiv \cup S'_i$ of open and closed sets under T' (i.e. each S'_i is open or closed in the topology T'). $f^{-1}(\cup S'_i) = \cup f^{-1}(S'_i)$ is a basic set-theory result. As a countable union of measurable sets, the right side is measurable in Σ.

Note that the converse is not true. There are more measurable functions than continuous ones. Consider $S \equiv f^{-1}(O')$ for $O' \in T'$. If f is a measurable fn, $S \in \Sigma$ since $O' \in \Sigma'$. However, not every measurable set is open — so it is possible for S not to be open.

This is just an example of proposition 4.3. In our case, the generating set Z' is the topology T'. But wait — we said the converse doesn't hold, yet in proposition 4.3 we have an "iff". What gives? The two cases are different. The "iff" speaks only to measurability. The equivalent condition to our converse would be that the inverse image of the subbasis Z' not only is measurable but is a particular subbasis Z for Σ. If f is measurable, it is *not* the case that (for Z' a subbasis of Σ') $f^{-1}(Z')$ is a subbasis for Σ. It is a subset of Σ, but need not constitute a subbasis for it — let alone any specific subbasis. In our case, T and T' are the relevant subbases. The inverse images of open sets are measurable, but need not be open.

4.3. **Borel and Radon measures.**

A **Borel measure** refers to a measure μ on the Borel algebra Σ of a topology T. However, the specifics of the definition vary. T may or may not be required to be Hausdorff and/or locally compact. If T is Hausdorff, μ may or may not be required to be locally finite or assign compact sets finite measure. We won't fuss over the term, but when it arises in the context of a theorem (ex. the Riesz representation theorem), we must take care that the choice is either clear or irrelevant.

Fortunately, the version of the Riesz representation theorem we will use involves a Hausdorff, locally compact space and explicitly states that μ is locally finite, so there won't be any ambiguity.

We'll often be sloppy with language, saying that the Borel measure is on T or the underlying set S or something similar. It is to be understood that we always mean that it is on the associated Borel algebra — unless otherwise stated (ex. if we are working in the completion).

Regular Borel Measure: Regardless of our definition of Borel measure, suppose we have a Borel algebra Σ for Hausdorff topological space T (with T having whatever other properties our definition of Borel measure requires). There may be many distinct Borel measures on Σ. A "regular" Borel measure is one in which we can approximate the measure of a set both by that of the open sets containing it and that of the compact sets contained in it.

Formally, given underlying set S, topology T on S, Borel algebra Σ for T, and measure μ on Σ, we require that for each $s \in \Sigma$, $\mu(s) = \inf_{O \in T}\{\mu(O); s \subseteq O\} = \sup_{K \in K(T)}\{\mu(K); K \subseteq s\}$, where $K(T)$ denotes the set of compact subsets of S under T (these need not be elements of T itself, of course). The compact supremum condition by itself is called "inner regularity" and the open infimum condition by itself is called "outer regularity".

The definition is actually a bit more general than we've stated. We can define the term "regular measure" in this manner for *any* σ-algebra containing the topology T, not just its Borel algebra. Of course, any such σ-algebra contains the Borel algebra.

We said "regardless of our definition of Borel measure", but the set of Borel measures on Σ and the set of regular Borel measures on Σ both depend on that choice. We just mean that the definition of "regularity" isn't confined to a specific choice. The only reason we must include Hausdorff in the definition is so that compact sets are measurable. Some older treatments omit this and use a supremum of closed sets rather than compact ones.

Radon measure: For T Hausdorff, a Borel measure that is both locally finite and inner regular.

If T is also locally compact, inner regularity implies outer regularity, and a Radon measure would be locally finite and regular. In this case, we may be tempted to imagine that the combination of inner regularity and the measure being finite on compact subsets implies that the Radon measure must be finite. We could argue that regularity forces every measurable set s to have finite measure because the supremum is over compact sets and all compact sets have finite measure. However, this is incorrect. Each $\mu(K)$ is finite, but their supremum need not be.

As mentioned, if our space is compact, then any locally finite Borel measure on it is finite. Unfortunately, the space of microstates we care about for the AEAP (i.e. Ω_E) is only locally compact rather than compact, so this won't help us.

The following two results tells us how Borel algebras (and the measures on them) change if we restrict ourselves to a topological subspace or a σ-subalgebra.

Prop 4.8: Let T be a topology on X, let $X' \subset X$ be an open subset, let T' be the subspace topology on X', and let Σ and Σ' be the Borel algebras of X and X'. Then (i) $\Sigma' \subseteq \Sigma$, (ii) any measure on Σ restricts to a measure on Σ', and (iii) if T is locally compact and μ is a Radon measure then so is its restriction.

Note that we are *not* saying that Σ' is a σ-subalgebra of Σ. It is a subset, and it is a σ-algebra on X', but it is *not* a σ-algebra on X. For one thing, it does not contain X itself.

This differs from proposition 4.1 in that we are not restricting to a σ-algebra on X.

Pf: (i) Since X' is open in T, the entire subspace topology $T' = \{O \cap X'; O \in T\}$ consists of open sets of T. I.e., it is a subset of T and thus contained in Σ'. This means that countable unions are in Σ' as well. The complement now is relative to X', not X. Given $O' \in T'$, $O' = X' \cap O$ for some $O \in T$. Therefore, $X' - O' = (X' \cap X) - (X' \cap O) = X' \cap (X - O)$, which is in Σ since X, X', and O are. Therefore, $\Sigma' \subseteq \Sigma$. (ii) Given measure μ on Σ, we can just restrict to $\Sigma' \subseteq \Sigma$. This trivially produces a measure on Σ' (as a σ-algebra on X', not X) because nonnegativity and countable additivity survive restriction. (iii) To admit a Radon measure, T must be Hausdorff. Any subspace of a Hausdorff space is Hausdorff, so T' is as well. Any open subspace of a locally compact space is locally compact, so T' must be as well. Let $i : X' \to X$ be the embedding map, and let K be a compact set in T'. A continuous function maps compact sets to compact sets, and i is continuous, so $i(K) = K$ is compact in T. I.e., any compact set of T' is compact in T. For a Hausdorff, locally compact space, local finiteness is equivalent to every compact set having finite measure. Denote by μ' the restriction of μ. By definition, $\mu(s) = \mu'(s)$ for every $s \in \Sigma'$. For a Hausdorff space, every compact set is closed, and thus in the Borel-algebra. Given a compact K relative to T', we know $K \in \Sigma'$ and thus in Σ and that it is compact relative to T. Therefore, $\mu(K) = \mu'(K)$, but $\mu(K)$ is finite since μ is locally finite. Now, consider some set $s \in \Sigma'$ and $\sup \mu(K)$ over all compact sets relative to T' that are subsets of s. Both s and each such K are in Σ', and all such K are compact relative to T as well. From the inner regularity of μ, we know that $\sup \mu(K) = \mu(s)$, where the supremum is over all compact subsets of s relative to T. However, any such $K \subseteq s$ is a subset of X' too, since $s \subseteq X'$, so the collections of compact K's in the two suprema are the same. I.e., we once again get $\mu(s) = \mu'(s)$. A Radon measure therefore restricts to a Radon measure in this case.

Prop 4.9: Let T be a topology on X, and let μ be a Radon measure on some σ-algebra Σ containing the Borel algebra of T. If $\Sigma' \subset \Sigma$ is a σ-subalgebra that also contains the Borel algebra, then μ restricts to a Radon measure on it.

Pf: We saw in proposition 4.1 that μ restricts to a measure. Local finiteness is preserved because all open sets are still in Σ'. If every $x \in S$ has an open neighborhood of finite measure, that open neighborhood is in both Σ and Σ'. Regularity also still holds. The number of measurable sets has decreased, but Σ' still contains all compact and open sets — and the restriction of μ keeps the same measures for them. Therefore, any suprema or infima are unchanged. We're simply demanding regularity of a smaller number of sets, so it remains intact.

4.4. **Completion of a measure.**

Although Borel algebras are often the focus of discussions about measures on topological spaces, they are not what we typically end up working with. For $\mathbb{R}^n$, we work with the Lebesgue measure, which is defined on a larger σ-algebra — the measure-theoretic completion of the Borel algebra.

This should not be confused with the notion of topological completion. In this section, "completion" will mean measure-theoretic.

Denote by (S, Σ, μ) the measure μ on σ-algebra Σ. The completion of (S, Σ, μ) is a specific (S, Σ', μ') s.t. $\Sigma \subseteq \Sigma'$ and $\mu'|_\Sigma = \mu$. I.e, Σ' extends Σ and μ' extends μ to Σ'.

There may be many such extensions, but the completion is a specific one that is constructed via a particular prescription. We won't go into the details of that construction here, instead confining ourselves to a general description of what it accomplishes. See [13], theorem 1.36 for details.

Σ confers a notion of measurability on certain subsets of the underlying set S. Under a particular μ, some of these may have measure zero. These are called "negligible" or "null" or "μ-null sets". If a set has measure zero, it would make sense to also assign measure zero to all of its subsets. However, some of those subsets may not be measurable in Σ. For lack of a better term, let's call a set "sub-null" (under μ) if it is a non-measurable subset of a null set of μ.

The basic idea of the completion is to include (and assign measure zero to) all sets which are "sub-null" under μ. I.e., we add in all the sets which should have measure zero but aren't measurable under Σ. Of course, it isn't quite that simple. To obtain a σ-algebra, we must also include complements of sub-null sets, countable unions of sub-null sets with already-measurable sets, etc. Formally, we generate Σ' from $\Sigma \cup Z$, where Z denotes the set of sub-null sets under μ. It is neither surprising nor difficult to show that the completion is unique.

Note that the completion depends on the measure involved. We complete (S, μ, Σ) to (S, μ', Σ'). We cannot speak of the "completion" of Σ in isolation. Even though Σ tells us which sets are measurable, we don't know which are measure zero (and thus which unmeasurable sets are "sub-null") until we are given μ.

By taking the completion, we expand the set of measurable sets as much as possible without materially changing anything. I.e., every newly measurable set differs from some existing measurable set in a measure-zero way. We've simply added all the sets whose measures we can trivially deduce from the measures of existing sets.

Prop 4.10: A finite measure completes to a finite measure, an infinite measure completes to an infinite measure, and a probability measure completes to a probability measure.

Pf: Completion doesn't change the overall volume, so $\mu'(S) = \mu(S)$.

Prop 4.11: Any set in the completion of (S, Σ, μ) can be written $s \cup n$ for some $s \in \Sigma$ and some sub-null set n s.t. $s \cap n = \emptyset$ (where we allow $n = \emptyset$).

Pf: (sketch): The completion is generated by $\Sigma \cup Z$, where Z is the set of sub-null sets of Σ under μ and also is taken to include $\emptyset$. Let $\cup n_i$ be a countable union of sub-null sets. Since each $n_i \subset s_i$ for some s_i with measure zero, $\cup n_i \subset \cup s_i$. However, μ is countably additive, so $\mu(\cup s_i) = \sum_i \mu(s_i) = \sum_i 0 = 0$. I.e. $\cup s_i$ is a measure-zero set, so any subset of it is sub-null. Therefore $\cup n_i \in Z$. We already know that Σ is closed under countable unions. Any countable union of elements of Σ and Z can be written $(\cup s_i) \cup (\cup n_j)$ for some countable set of $s_i \in \Sigma$ and some countable set of $n_j \in Z$. We've shown that each of these two countable unions is a member of its respective class — so we can write the overall union as $s \cup n$, where $s = \cup s_i \in \Sigma$ and $n = \cup n_j \in Z$. Next, consider complements. We know that $\overline{s} \in \Sigma$, and is of the requisite form $s \cup \emptyset$. It therefore suffices to show that $\overline{n}$ is of the form $s' \cup n'$ for some $s' \in \Sigma$ and some $n' \in Z$. As a sub-null set, $n \subset s_0$ for some measure-zero set s_0. Define $w = s_0 - n$. Since $w \subset s_0$, it too is a sub-null set. I.e., s_0 is the disjoint union of sub-null sets $n \cup w$. We therefore have a partition of S: $S = \overline{s_0} \cup n \cup w$, where $\overline{s_0}$, n, and w are pairwise disjoint. This means that $\overline{n} = \overline{s_0} \cup w$. Since $\overline{s_0} \in \Sigma$ and $w \in Z$, we have our decomposition.

Prop 4.12: For a finite measure μ, the operations of normalization to a probability measure and completion commute.

I.e., we get the same P' if we (a) complete μ to μ' and then normalize the completion to a probability measure P' or (b) normalize μ to a probability measure P and then take the completion of P to P'.

Pf: (sketch): Since the overall volume doesn't change, the divisor $\mu'(S) = \mu(S)$ is the same. Let's call it V. The set of sub-null sets doesn't change under normalization, so the completion σ-algebra Σ' is the same whether we normalize first or not. Let (i) (Σ, μ) be the original finite measure, (ii) (Σ, P_μ) be the normalized original measure, (iii) (Σ', μ') be the completion of (Σ, μ), (iv) $(\Sigma', P_{\mu'})$ be the normalization of (Σ', μ'), and (v) (Σ', P'_μ) be the completion of (Σ, P_μ). Our goal is to show that $P'_\mu = P_{\mu'}$. On $\Sigma \subseteq \Sigma'$, $\mu'|_\Sigma = \mu$, so $(\mu'|_\Sigma)/V = \mu/V$. The right side is P_μ and the left side is just $P_{\mu'}|_\Sigma$. I.e., $P_{\mu'}|_\Sigma = P_\mu$. The completion doesn't change the measure of existing sets and only adds sets of measure zero. Since it is unique and since the same σ-algebra Σ' is involved in both completions, $(\Sigma', P_{\mu'})$ must be the completion of (Σ, P_μ). I.e., we must have $P_{\mu'} = P'_\mu$.

We need to be a bit careful interpreting zero probabilities, because $P(s) = 0$ in probability theory does not mean that s is impossible. However, every newly minted measure-zero set (i.e. formerly sub-null set) is a subset of an existing measure-zero set and thus subject to (at worst) the same considerations as its parent set in this regard. We're doing nothing but expanding in the obvious way the sets we can talk about.

Any given (S, Σ, μ) is either complete or not from a measure-theoretic standpoint. I.e., completeness is a property of a measure. If a measure is complete, then it is its own completion.

As the following proposition tells us, measurability of a function is preserved by a completion (under any μ) of the source Σ. However, it is not preserved by a completion (under any μ') of the destination Σ'.

Prop 4.13: If f is measurable relative to (S, Σ) and (S', Σ'), then it is measurable relative to *any* completion of (S, Σ) as well.

What we mean by this is that given any μ on Σ , f is measurable relative to the completion Σ_C (and Σ').

Pf: Any completion of Σ contains Σ, so proposition 4.2 gives us the result.

This is *not* the case for completions of (S', Σ'). Suppose we have some μ' on Σ', for which the completion is (Σ'_C, μ'_C). As mentioned in proposition 4.11, any set in the completion Σ'_C can be written $s' \cup n'$ for some $s' \in \Sigma'$ and some sub-null set n'. $f^{-1}(s' \cup n') = f^{-1}(s') \cup f^{-1}(n')$. We know that $f^{-1}(s') \in \Sigma$, since f is measurable against Σ and Σ'. However, there is no reason that $f^{-1}(n')$ should be in Σ. In general, it is not. f is measurable against the completion Σ'_C iff $\Sigma'_C \subseteq \Sigma'_f$ (using our earlier notation).

Just as completion commutes with normalization for finite measures, completion commutes with restriction in the sense of proposition 4.8. The following result is general, but applies to the Borel algebra case as well.

Prop 4.14: Given (S, Σ), let $S' \subset S$, and let $\Sigma' \subset \Sigma$ be a σ-algebra on S'. For any measure μ on Σ, we can restrict μ to a measure on Σ' in the obvious way. The operations of completion and restriction commute.

This mirrors our disussion from proposition 4.8. As is the case there, Σ' is *not* a σ-algebra on S (it doesn't even contain S). We're just requiring it to be a subset of Σ that happens to be a σ-algebra on S'.

The completion of the restriction is unambiguous, but we have to be a little careful what we mean by the restriction of the completion. We haven't specified which σ-algebra on S' we are restricting to. Obviously, we'll mean that of the "completion of the restriction". By definition, this will be a σ-algebra on S', but we must show that it is a subset of the completion of (Σ, μ). Otherwise, we can't restrict to it.

Pf: The restriction of μ to Σ' preserves nonnegativity and countable additivity, so μ is a measure on Σ'. Let (i) (Σ_C, μ_C) be the completion of (Σ, μ), (ii) (Σ', μ') be the restriction of μ to Σ', and (iii) (Σ'_C, μ'_C) be the completion of (Σ', μ'). Any sub-null set of (Σ', μ') is a sub-null set of (Σ, μ), so the completion of the former must be a subset of the completion of the latter. This addresses the issue of whether the "completion of the restriction" is a subset of the completion of (Σ, μ). I.e., we are guaranteed that $\Sigma'_C \subseteq \Sigma_C$, and we may legitimately define (iv) (Σ''_C, μ''_C) to be the restriction of (Σ_C, μ_C) to Σ'_C. Our goal is to show that $\Sigma''_C = \Sigma'_C$ and $\mu''_C = \mu'_C$. Since all the μ variants assign 0 to sub-null sets and agree on any common domain, we need only show that $\Sigma''_C = \Sigma'_C$. Since Σ''_C is the restriction of Σ_C to Σ'_C, it must be $\subseteq \Sigma'_C$. Consider $s \in \Sigma'_C$. By proposition 4.11, $s = s_1 \cup n$ for some $s_1 \in \Sigma'$ and some sub-null set n of Σ' under μ. However, every element of Σ' is in Σ and every sub-null set of Σ' is a sub-null set of Σ. I.e., $s_1 \cup n \in \Sigma_C$. We are restricting to Σ'_C, so $s_1 \cup n$ survives the restriction and just ends up in Σ''_C. I.e., $\Sigma'_C \subseteq \Sigma''_C$ as well, and the two must be equal.

Prop 4.15: Let T be a locally compact topology on set X, and let μ be a Radon measure on its Borel algebra Σ. Then the completion (X, Σ', μ') of (X, Σ, μ) is also a Radon measure.

Pf: The definition of Radon measure requires T to be Hausdorff, so it is that as well. For a locally compact, Hausdorff space, local finiteness of a measure (which is defined on any σ-algebra containing the Borel algebra) is equivalent to "every compact set has finite measure". As an extension of μ, the completion satisfies $\mu'(s) = \mu(s)$ for every $s \in \Sigma$. Taking the completion doesn't affect the underlying topology or notion of compactness, so $\mu'(K) = \mu(K)$ for every compact set K under T, and $\mu(K)$ is finite since μ is a Radon measure. We therefore have local finiteness of μ'. By the same token, for any $s \in \Sigma$ the supremum is over the same set of compact sets (with the same measures) as before. So $\mu'(s) = \mu(s) = \sup_{K \subset s} \mu(s) = \sup_{K \subset s} \mu'(s)$ (with K denoting compact subsets as usual). Now, consider some set in $s' \in (\Sigma' - \Sigma)$. By proposition 4.11, any set in the completion can be written $s' = s \cup n$ for some disjoint pair s and n. By countable additivity, $\mu(s') = \mu'(s) + \mu'(n) = \mu(s) + 0$. Any $K \subset s$ also satisfies $K \subset s \cup n$, so the supremum for s' is over at least as many K's as for s. We could show there are no compact subsets of $s \cup n$ that aren't subsets of s, but this is unnecessary. We merely need to observe that by monotonicity, every such $\mu'(K) \leq \mu'(s) = \mu(s)$. However, inner regularity of μ means that the supremum already equals $\mu(s)$. Even if we were to include more K's, they would have to be $\leq$ this and wouldn't affect the supremum. Therefore, we have inner regularity for μ' as well.

In summary, it doesn't matter whether we complete before or after restriction or normalization, and completion preserves a Radon measure.

By monotonicity, if an extension of (Σ, μ) contains sub-null sets of Σ under μ, it must assign them zero measure. An extension may or may not contain such sets, but there is no flexibility in the measure assigned to them. It follows that any extension that is not the completion either must (i) fail to include some sub-null sets or (ii) include some non-sub-null sets. The latter can be assigned zero measure in their own right, but need not. Obviously, any extension which contains the completion must also be an extension *of* the completion.

At first glance, it may seem that the completion of (Σ, μ) is terminal, and no other extension can be complete. However, this is not the case. An extension to a larger σ-algebra can introduce new null sets, thus engendering new sub-null sets. It is quite possible to have complete measures that contain other complete measures.

We'll continue to focus on Borel algebras, but almost everything we do with them works just as well for their completions (and we'll point out when it does not). When we define the notion of a "uniform" probability measure, we'll be able to speak of its completion as uniform too. No new volume has been added, so no change is necessary. If we have a natural choice of probability measure on the Borel algebra, we have a corresponding (unique) completion of that probability measure.

We could work directly with the completion instead of the Borel algebra, but this is inconvenient for our purposes because it presumes knowledge of the relevant completion (i.e. the relevant μ). The Borel algebra, by contrast, depends solely of the topology. Note that *for a given Σ* there is a bijection between measures (Σ, μ) and their completions (Σ', μ'). However, it is quite possible for (Σ_1, μ_1) and (Σ_2, μ_2) to have the same completion (Σ', μ'). The completion has a unique restriction to a given Σ, but can restrict to different Σ's. For example, suppose we start with (Σ, μ). Let Y be the set of measure-zero sets of Σ under μ, and let Z be the set of sub-null sets as usual. Suppose Y_1 and Y_2 are subsets of Y s.t. each contains sets that aren't subsets of any set in the other (i.e. $\exists y_1 \in Y_1$ s.t. y_1 isn't a subset of any element of Y_2, and vice versa). Let $Z_1 \subset Z$ and $Z_2 \subset Z$ contain the sub-null sets that are subsets of elements in Y_1 and Y_2. Then the σ-algebras generated by $\Sigma \cup Z_1$ and $\Sigma \cup Z_2$ are distinct but have the same completion: the σ-algebra generated by $\Sigma \cup Z$.

5. Uniform probabilities

Let's put aside the AEAP for the moment and consider the general question of what is meant by "uniform" probabilities for an infinite set. Formal probability theory traffics in probability fields, and we will take that approach.

We're not doing so just as a preference, but because the machinery of formal probability theory is essential to the analysis of probabilities on infinite sets. In fact, dealing with infinite sets is one reason the measure-theoretic framework for probability was developed.

5.1. **Uniform probability field.**

What does it mean to have a uniform probability measure on (S, Σ)? Absolutely nothing. For a general (S, Σ), we need additional information or structure in order to define a notion of uniform probabilities. Let's think about what this entails.

Intuitively, we want the probability of an event to be proportional to its size. However, "size" can't mean cardinality since the cardinalities can be infinite here. What we mean by size must be some sort of notion of volume. We're dealing with the sets in a σ-algebra, so we have a natural associated notion of "measurable". Each choice of measure μ on Σ assigns a volume to every measurable set. Our intuitive notion of "uniform" is that the relative probabilities of events should be proportional to their volumes.

Formally, suppose we have measure μ on (S, Σ) and we consider two sets $A, B \in \Sigma$, both with nonzero measures under μ (i.e. $\mu(A) > 0$ and $\mu(B) > 0$, which also implies that they are nonempty). Our intuitive notion of "uniform" probability (under μ) is that their relative probabilities should be $\mu(A)/\mu(B)$. This is effected by defining the probability field $P(s) \equiv \mu(s)/\mu(S)$ for all $s \in \Sigma$. However, there is a hitch. Doing so requires that $\mu(S) < \infty$. I.e., we need μ to be a finite measure. If it is not, we can't assign uniform probabilities — for the same reason we can't assign them to the integers. Note that we can always speak of the relative probabilities of two sets of finite volume (i.e. $\mu(A)/\mu(B)$), but we cannot turn this into a probability measure unless $\mu(S) < \infty$.

We thus want a choice of probability measure — but which one? *Every* choice of probability measure gives rise to a corresponding definition of "uniform" probabilities. We therefore need some sort of additional information, constraints, or structure that reduces the choice to a single probability measure that we deem "natural".

It suffices for our additional information to furnish a finite measure, since every finite measure normalizes to a probability measure.

Absent any such additional information, constraints, or structure, there is no such thing as a "uniform" probability field. More precisely, there is no mechanism to designate any one of the many probability fields on Σ as "uniform".

There is a class of cases in which a natural definition does exist, however. If (S, Σ) has a partition basis B, the elements of B are "atomic" from the standpoint of probability theory. As mentioned earlier, any measure is uniquely determined by its values on the basis elements in that case. We therefore have a natural concept of "uniform" probabilities: we just assign the basis sets equal probabilities. However, it only is possible to do so when B is finite. Just as we cannot impose a uniform probability distribution on the integers, we cannot do so for an infinite B. I.e. we have such an intrinsic definition of "uniform" iff a finite partition basis exists. This does not require the underlying set S to be finite, but it does mean that Σ is finite. Since every finite Σ has a partition basis, the converse holds as well. Note that if Σ is (or contains) the Borel algebra of a topology, as in our case, a finite Σ implies a finite topology. I.e., the situation where a partition basis offers an intrinsic definition of uniformity only arises in simple or contrived circumstances. It certainly does not apply to Ω or Ω_E, which both have infinite topologies.

In summary, to impose a meaningful notion of uniform probabilities on our Borel Algebra, we require the specific application at hand to furnish a unique "natural" choice of probability measure on it. We then can employ the notion of "uniform probabilities" conferred by this probability measure.

5.2. **Aside: Uniform random variable.**

We also can speak of "uniform" probabilities for a random variable X. Recall that a **random variable** on (S, Σ) is a function $X : S \to \mathbb{R}$ s.t. $X^{-1}((-\infty, y)) \in \Sigma$ for every $y \in \mathbb{R}$. Equivalently, it is a real function that is measurable relative to Σ and the Borel algebra Σ_{Bor} of $\mathbb{R}$.

Proposition 4.4 tells us that the set $\{(-\infty, y); y \in \mathbb{R}\}$ is a subbasis for Σ_{Bor}, and proposition 4.3 tells us that $f^{-1}(s) \in \Sigma$ for all s in a subbasis of Σ_{Bor} iff f is measurable relative to Σ_{Bor}. Older treatments employ the interval definition of a random variable, and modern treatments favor phrasing things in terms of measurability relative to Σ_{Bor}.

Measurability is an important constraint on f, even in the simplest cases. For example, consider $S = \{a, b, c\}$ and $\Sigma = \{\emptyset, (a), (b, c), S\}$ and $f : (a, b, c) \to (1, 1, 2)$. In this case, f is not a measurable function because $f^{-1}((0.5, 1.5)) = (a, b)$ is not a measurable set in Σ.

Given (S, Σ), a random variable X, and a probability measure P on Σ, we can derive a cumulative distribution function (CDF) for X. This is a monotonically nondecreasing function $F : \mathbb{R} \to [0, 1]$ defined as $F(y) \equiv P(X^{-1}((-\infty, y)))$. If F is everywhere-differentiable, we can define a corresponding probability density function (PDF) as its derivative $\rho(y) \equiv F'|_y$. For any given P, every random variable has an associated CDF, but not necessarily a PDF.

Note that we can extend all these definitions to $X : S \to \mathbb{R}^n$ easily enough.

Technically, a PDF requires that the CDF be differentiable everywhere except a measure-zero set of points (under the usual measure on $\mathbb{R}$).

For our present purpose, random variables are not a useful vehicle. $S = \Omega_E$, and Σ is its Borel algebra (or the completion of that Borel algebra under whatever measure we choose) . The AEAP requires "uniform probabilities" on S itself, not for some random variable. The best embodiment of this notion is not a CDF for a function $S \to \mathbb{R}$, but rather a measure on Σ — which gives us a definition of "volume" and thus of "equal volumes".

Note that a choice of finite μ, although it defines a notion of "volume", does *not* beget a random variable on Σ. It is a function $\Sigma \to \mathbb{R}$, not a function $S \to \mathbb{R}$. A random variable is a labeling of elementary events, which is not what we need here. A given CDF *is* a random variable in its own right, but this doesn't help us either — since that CDF (as such) arises from an existing measure and random variable.

Although we won't make use of the notion of uniform probabilities for a random variable, let's nonetheless consider what it means. We'll begin with a couple of useful results.

Prop 5.1: Given (S, Σ) and any $f : S \to S'$, we have an induced σ-algebra Σ'_f on S' consisting of all sets $s' \subseteq S'$ s.t. $f^{-1}(s')$ is measurable. f is measurable relative to Σ and Σ'_f.

Pf: $f^{-1}(S') = S$ and $f^{-1}(\emptyset) = \emptyset$, so $S, \emptyset \in \Sigma'_f$. Given $s' \in \Sigma'_f$, let $s = f^{-1}(s')$. $f^{-1}(\overline{s'}) = \overline{f^{-1}(s')} = \overline{s}$, which is in Σ, so $\overline{s'} \in \Sigma'_f$. Suppose we have a countable collection of events $s'_i \in \Sigma'_f$. Since $f^{-1}(\cup s'_i) = \cup f^{-1}(s'_i) \in \Sigma$, we have $\cup s'_i \in \Sigma'_f$. By the definition of Σ'_f, f is measurable.

Note that Σ'_f depends on the choice of Σ, even though our notation obscures this.

Σ'_f is the largest σ-algebra on S' for which f is measurable against Σ and Σ'_f. As we show below, it contains all other such σ-algebras.

Prop 5.2: Given (S, Σ) and (S', Σ') and any measurable (relative to Σ and Σ') function $f : S \to S'$, Σ' is a σ-subalgebra of the induced Σ'_f.

Pf: By definition, Σ'_f contains every $s' \subseteq S'$ s.t. $f^{-1}(s') \in \Sigma$. Since f is measurable relative to Σ and Σ', any $s' \in \Sigma'$ has $f^{-1}(s') \in \Sigma$ and thus is an element of Σ'_f too. Since $\Sigma' \subseteq \Sigma'_f$ and both are σ-algebras, Σ' is a σ-subalgebra of Σ'_f.

I.e., any Σ' is a σ-subalgebra of every Σ'_f for which f is measurable against Σ and Σ'.

Suppose we have (S, Σ) and random variable $f : S \to \mathbb{R}$. Denote by $(\mathbb{R}, \Sigma_{\mathbb{R}})$, the standard σ-algebra on $\mathbb{R}$ and by $\mu_{\mathbb{R}}$ the standard measure on $\mathbb{R}$.

These could be defined as the Borel algebra and Borel measure on $\mathbb{R}$ or their completion (aka the Lebesgue measure and its σ-algebra). The choice won't make a difference for the present discussion. For convenience, we'll refer to elements of whichever we choose as "Borel sets", even though that term should — strictly speaking — be reserved for elements of the Borel algebra.

This f and Σ induce a Σ'_f on $\mathbb{R}$ via proposition 5.1. Since f is measurable relative to $\Sigma_{\mathbb{R}}$, proposition 5.2 tells us that $\Sigma_{\mathbb{R}} \subseteq \Sigma'_f$. This guarantees that every Borel set is measurable in Σ'_f as well.

If $f(S)$ is a finite (and thus discrete) set, our intuitive notion of "uniform probabilities" assigns equal probabilities to the elements of $f(S)$. This differs from uniform probabilities on a finite S in that we are assigning equal probabilities to values of f rather than points of S. That gives us a clear idea of how to proceed in the general case.

We are working with values of f rather than elements of S, and those value are in $\mathbb{R}$. We therefore expect the notion of "uniform probabilities" for a random variable to be conferred by the measure $\mu_{\mathbb{R}}$. Specifically, we expect that any two measurable sets $A_1, A_2 \subset \mathbb{R}$ of values should have relative probabilities $\mu_{\mathbb{R}}(A_1)/\mu_{\mathbb{R}}(A_2)$.

The term "measurable sets" deserves clarification. The largest σ-algebra against which f is measurable on that end is Σ'_f. We wish to at least be able to speak of the probability of an arbitrary interval of values for f. Such intervals generate $\Sigma_{\mathbb{R}}$, so we at least need the Borel algebra. For our purposes, we'll stick with A_1 and A_2 being measurable under the Borel algebra. This is the weakest possible constraint consistent with our needs.

There are a few problems with this approach as it stands. First, A_1 and A_2 must have finite measures (and A_2 must have nonzero measure, to allow the ratio). We can only assign relative probabilities, since $\mu_{\mathbb{R}}(\mathbb{R}) = \infty$. We also are assuming that A_1 and A_2 both reside in $f(S)$. Otherwise, our accounting of volume is off, and we're assigning equal probability to values that f can never take.

We do not care whether f is injective, however. We're assigning equal probabilities to values of f, not points of S. If many points of S map to the same value, it shouldn't affect the probability of the value under such a scheme.

To address these issues, we must modify our naive approach. First, we need to confine ourselves to the image of f. To be able to assign absolute probabilities to sets of values of f, we'll require that $f(S)$ have finite, nonzero measure. I.e., we need $0 < \mu_{\mathbb{R}}(f(S)) < \infty$. We'll also require that, aside from a measure-zero set of points (under $\mu_{\mathbb{R}}$, of course), $f(S)$ is a countable union of disjoint intervals (open or closed or half-open — it doesn't matter) in $\mathbb{R}$.

It turns out that this last requirement doesn't impose a meaningful constraint (at least for Σ_{Bor}). Although proposition 4.4 tells us that such intervals form a subbasis of Σ_{Bor}, they do not form a basis — which is what would be needed here. It is possible to show that every open set in $\mathbb{R}$ is a countable union of open intervals. However, there are sets in Σ_{Bor} which cannot be expressed as a countable union of intervals of any sort. These include $\mathbb{R} - \mathbb{Q}$ and the Cantor set. However, it is possible to show that every Borel set differs from a countable union of intervals by a set of measure-zero. [Basically, a closed set on $\mathbb{R}$ differs from an open set by at most two points. Closure under countable unions and complements mean that we can only ever deviate from an open set by a countable number of points. Any open set in $\mathbb{R}$ is a countable union of open intervals (a standard analysis result), so we differ from a countable union of intervals by a countable set of points. In the standard measure on $\mathbb{R}$, any countable set is measurable with measure zero.] Since this is precisely the requirement we've imposed, it doesn't constrain away any Borel sets. Of course, we have no guarantee that $f(S)$ itself is a Borel set. However, counterexamples tend to be pathological and contrived. In practice, this assumption doesn't limit us much. Note that $f(S)$ *is* in Σ'_f, since $f^{-1}(f(S)) = S$.

Let's construct a CDF for f which reflects our intuition about uniform probabilities. Let $f(S) \cup Y_1 = Y_2 \cup (\cup_i A_i)$ be the expansion of $f(S)$ in disjoint intervals, where Y_1 is a measure zero set of holes in $f(S)$ and Y_2 is a measure zero set of non-interval points of $f(S)$.

We'll begin by defining a PDF ρ. Define $\rho(x) = 0$ outside the intervals (i.e. for all $x \notin \cup_i A_i$) and $\rho(x) = 1/\mu_{\mathbb{R}}(f(S))$ for $x \in \cup_i A_i$. We can integrate this to get a staircase-shaped CDF: $F(x) \equiv \int_{-\infty}^{x} \rho(x')dx'$.

Via one of the core results of probability theory, a CDF for a random variable defines a unique probability measure on Σ_{Bor}. For the uniform CDF defined above, this probability measure P has value $P(A_i) = \mu_{\mathbb{R}}(A_i)/\mu_{\mathbb{R}}(f(S))$ on the intervals themselves and extends to the whole Borel algebra in a natural way.

The core result follows because (i) we can obtain a premeasure on the intervals via $P([x, y)) = F(y) - F(x)$, (ii) it is easy to show that this extends to a measure on Σ_{Bor}, and (iii) this measure is unique. For uniqueness of this extension see, for example, [12], theorem 3.7.

Formally, we define $P(s) \equiv \mu_{\mathbb{R}}(s \cap (\cup_i A_i))/\mu_{\mathbb{R}}(f(S))$ for every $s \in \Sigma_{Bor}$. This works for both the Borel algebra and its completion under $\mu_{\mathbb{R}}$. To see that it is a probability measure, we observe that: (i) $P(\mathbb{R}) = \mu_{\mathbb{R}}(\cup A_i)/\mu_{\mathbb{R}}(f(S)) = 1$ since $\cup A_i$ differs from $f(S)$ by a set of measure 0 under $\mu_{\mathbb{R}}$. (ii) For disjoint countable collection $\{s_i\}$, $P(\cup s_i) = \mu_{\mathbb{R}}((\cup s_i) \cap (\cup A_i))/\mu_{\mathbb{R}}(f(S)) = \mu_{\mathbb{R}}(\cup(s_i \cap (\cup A_i)))/\mu_{\mathbb{R}}(f(S))$. The right side is a disjoint union, so by countable additivity of $\mu_{\mathbb{R}}$ we get a sum of $P(s_i)$, and P satisfies countable additivity. (iii) $P(s)$ is patently nonnegative.

Our uniform CDF therefore gives us a probability measure on $(\mathbb{R}, \Sigma_{Bor})$, embodying the notion of "uniform probabilities" for the random variable f.

Prop 5.3: Given (S, Σ, μ), set S', and any function $f : S \to S'$, there is a natural push-forward of μ to a measure (S', Σ'_f, μ'_f). It is given by $\mu'(s') \equiv \mu(f^{-1}(s'))$. A probability measure pushes forward to a probability measure.

Pf: μ'_f is patently nonnegative. Consider a countable collection of disjoint events $s'_i \in \Sigma'_f$. Let $s_i \equiv f^{-1}(s'_i)$. By the definition of Σ'_f, $s_i \in \Sigma$, so $\cup s_i$ is as well. $f^{-1}(\cup s'_i) = \cup f^{-1}(s'_i)$, and if $s'_i \cap s'_j = \emptyset$, $f^{-1}(s'_i) \cap f^{-1}(s'_j) = \emptyset$. We therefore have a countable collection of disjoint sets $s_i \in \Sigma$. By the countable additivity of μ, $\mu(\cup s_i) = \sum \mu(s_i)$. So $\mu(f^{-1}(\cup s'_i)) = \mu(\cup f^{-1}(s'_i)) = \mu(\cup s_i) = \sum \mu(s_i) = \sum \mu'_f(s'_i)$, and we have countable additivity for μ'_f. Since $\mu'_f(S') = \mu(f^{-1}(S')) = \mu(S)$, both have the same total volume. If μ is a probability measure, so is μ'_f.

Although there is a one-to-one correspondence between CDFs for f and probability measures on Σ_{Bor}, there is no such relationship between CDFs for f and probability measures on Σ. A given probability measure on Σ pushes forward to a probability measure on Σ'_f, which restricts to a probability measure on Σ_{Bor}. It therefore yields a CDF for f. However, given a CDF for f, there may be no measure, one measure, or multiple measures on Σ that produce it in this manner. Consider $S = \{a, b, c, d\}$. As an example with multiple measures, let $\Sigma = 2^S$ and $f : (a, b, c, d) \to (1, 1, 2, 2)$. A CDF has 1 degree of freedom. We specify a choice of $p(1)$, let $p(2) = 1 - p(1)$, and build our CDF from these. However, a probability measure has 3 degrees of freedom: any set of nonnegative values on the partition basis $(\{a\}, \{b\}, \{c\}, \{d\})$ that sum to 1. There are more probability measures than CDFs in this case. As an example with none, use the same S, let $\Sigma = \{\emptyset, (a, b), (c, d), S\}$, and let $f : (a, b, c, d) \to (1, 2, 3, 4)$. Now, a CDF for f has 3 degrees of freedom and a a probability measure on Σ has only 1 degree of freedom (since the partition basis has only two elements). There are more CDFs than probability measures in this case.

Any μ on Σ pushes forward to a μ'_f on Σ'_f, which then restricts to a μ'_f on $\Sigma_{\mathbb{R}}$. Unfortunately, the opposite does not hold. We cannot pull back a measure on Σ'_f (or on $\Sigma_{\mathbb{R}}$) to one on Σ.

This is evident from the discussion of CDFs above, but let's look at it from a σ-algebra standpoint. Suppose we have μ' on Σ'_f. Given $s \in \Sigma$, we are not guaranteed that $f(s) \in \Sigma'_f$, so we cannot use $\mu'(f(s))$. We could try something like the infimum of μ' over all open subsets of $f(S)$ containing $f(s)$ or the supremum over all compact subsets of $f(s)$, but this approach doesn't necessarily yield a measure on Σ (since the infima and suprema are in terms of $f(s)$, not s itself).

The probability measure on Σ_{Bor} induced by our uniform CDF for f is not special from the standpoint of any of the discussion above. There may be no, one, or many probability measures on Σ which push forward to it.

We therefore must content ourselves with the notion of a uniform CDF for f, and the probability measure this induces on Σ_{Bor}. Existence and uniqueness are not guaranteed for a corresponding probability measure on Σ.

6. Measure induced by a volume form on a manifold

Before we can establish that our measure on Ω_E is finite, we need a measure to begin with. We have a manifold $\Omega = T^*Q$, and we want to find a way to construct a natural finite measure on each $\Omega_E \subset T^*Q$. It turns out that any volume form on a manifold induces a corresponding measure on its Borel algebra. Finding a natural volume form therefore gives us a possible avenue to obtaining such a measure. As it happens, there *is* a natural volume form on Ω and Ω_E. We'll derive this shortly, but first let's see how a volume form induces a measure.

Suppose we have an orientable manifold M. Since it is orientable, it admits volume forms. As a topological space, it has an associated Borel algebra. Any volume form on M gives rise to a unique measure with the properties we need on this Borel algebra. The intuition for this is relatively straightforward, but the details are a bit involved.

Let ν be a volume form on M. As discussed earlier, we can compute the differential-geometric integral $\int_U f|_U \nu|_U$ for any open set $U \subseteq M$ and any suitable function $f : M \to \mathbb{R}$. With $f = 1$, we obtain the volume of U.

Let T denote the topology underlying M. Define a function $\mu : T \to \mathbb{R}^*$ via $\mu(U) \equiv \int_U \nu_U$ (i.e. $\int_U i_U^* \nu$). This μ takes each open set to its volume under ν and has two key features:

> Recall that R^* refers to the extended real line: $\mathbb{R} \cup \{-\infty, +\infty\}$. We define our functions to produce values in $\mathbb{R}$, but it is quite possible for the integral to be $\pm\infty$.

(i) $\mu(U) > 0$ for $U \neq \emptyset$ (and $\mu(\emptyset) = 0$) since we're integrating a volume form.

(ii) μ is countably additive.

> When we say "disjoint" (here and elsewhere), we mean pairwise disjoint. Pairwise disjoint implies mutually disjoint (i.e. $U_i \cap U_j = \emptyset$ for all i, j implies $\cap_i U_i = \emptyset$), but the converse need not hold. I.e. pairwise disjointness is the stricter condition.

> Pf: Let $U \equiv \cup U_i$. Since T is closed under arbitrary unions, we have $U \in T$. Therefore, $\mu(\cup U_i)$ is well-defined. The set $\{U_i\}$ forms a disjoint open cover of U. Therefore, for each i, the pair U_i and $\cup_{j \neq i} U_i$ form a two-set disjoint open cover of U as well. The only way two disjoint open sets can cover topological space U is if each is a union of connected components of U (not to be confused with the connected components of M itself). This means that each U_i is a union of connected components of U. The differential-geometric definition of integration then tells us that $\int_U \nu|_U = \sum_i \int_{U_i} \nu|_{U_i}$, a sum over the individual integrals.

If T were a σ-algebra, these two properties would qualify μ as a measure on it. As it is, we can think of μ as a sort of "premeasure", for lack of a better term. We wish to extend this premeasure on T to an actual measure on some σ-algebra containing T. The smallest possible σ-algebra containing T is its Borel algebra, so we'll at least need to extend μ to this. Intuitively, such an extension involves assigning a measure to closed sets and countable intersections.

> A σ-algebra is closed under complements and countable intersections, neither of which the topology is closed under. Therefore, we must assign measures to these new sets and any that are generated from them. If $\mu(M)$ is finite, we can define $\mu(C) \equiv \mu(M) - \mu(\overline{C})$, but the countable intersections are a bit trickier.

We're only interested in existence and uniqueness here, so we won't delve into the details of this particular construction. Existence and uniqueness are guaranteed by something called the "Riesz representation theorem."

> There are several Riesz representation theorems. The one we're referring to is known as the Riesz-Markov-Kakutani representation theorem.

6.1. **Premeasure as a functional.**

First, let's consider what we have so far. Given our manifold M and volume form ν, we constructed a function $\mu : T \to \mathbb{R}$ which is positive (on nonempty sets) and countably additive. However, μ isn't a measure because it is defined on a topology rather than a σ-algebra.

Let's now take a step back. Rather than explicitly trying to extend μ to the Borel algebra, we'll focus on the volume form ν on M. Consider the set $C^0(M)$ of functions $M \to \mathbb{R}$ that are continuous with respect to the topology of M and the standard topology on $\mathbb{R}$. It is easy to see that $C^0(M)$ is a vector space over $\mathbb{R}$. Loosely speaking, we can view ν as a linear functional that takes continuous functions on M to their integrals.

We say "loosely" because ν is *not* in fact linear on $C^0(M)$. We'll soon see why.

See our earlier discussion of how we can integrate a *continuous* function on a smooth manifold using a smooth volume form ν.

For any given $f \in C^0(M)$, $f\nu$ is an n-form, and we can compute $\int_M f\nu$. This defines a map $\hat{\mu} : C^0(M) \to \mathbb{R}^*$, which takes each continuous function $f : M \to \mathbb{R}$ to its integral $\hat{\mu}(f) \equiv \int_M f\nu$.

The image is $\mathbb{R}^*$ because the integral may be $\pm\infty$ in general. This is not just a minor technical issue; it is the central obstruction to $\hat{\mu}$ being a linear functional on $C^0(M)$ (and various other subspaces).

We cannot call $\hat{\mu}$ a linear functional because it is a map to $\mathbb{R}^*$ rather than $\mathbb{R}$.

This problem is more than mere semantics. $\mathbb{R}^*$ is not a vector space, or even an abelian group. $\infty + \infty = \infty$, which is impossible for an abelian group since $\infty + 0 = \infty$, so ∞ would have two additive inverses. Similarly, $\infty - \infty$ (or $\infty + (-\infty)$) is ill-defined. If we set it to 0, then $\infty + \infty - \infty$ is ∞ or 0 depending on the order of operations. I.e., we lose associativity. We also have issues with scalar multiplication of 0 and $\pm\infty$. Because $\hat{\mu}$ is not a map to a vector space, we can't call it linear.

However, $\hat{\mu}$ *is* a linear functional when restricted to any vector space of functions whose integrals are finite. The most commonly employed such subspaces are $L^1(M)$ and $L^2(M)$. We're considering continuous functions, so we would want something like $L^1(M) \cap C^0(M)$, the continuous absolute-integrable functions. Restricted to this, or any vector subspace of it, $\hat{\mu}$ is a linear functional. In our case, the subspace $C_C(M) \subset L^1(M) \cap C^0(M)$ (which we'll define momentarily) will be of interest, since it suffices for the Riesz representation theorem.

Although the notation obscures it, $L^1(M)$, $L^2(M)$, etc, depend on our choice of volume form ν, since that determines the integrals and thus the set of functions with finite integrals. However, $C^0(M)$ solely depends on the topology of M.

Both common forms of the Riesz-Markov-Kakutani representation theorem traffic in subspaces of $L^1(M) \cap C^0(M)$. One formulation requires a linear functional on $C_C(M)$ (the compactly supported continuous functions), and the other requires a linear functional on $C_0(M)$ (not to be confused with $C^0(M)$), the continuous functions that "vanish at infinity".

Given locally compact topological space X and field $F = \mathbb{R}$ or $F = \mathbb{C}$, a function $f : X \to F$ "vanishes at infinity" if for every $\epsilon > 0$, there exists some compact $K \subseteq X$ s.t. $|f(x)| < \epsilon$ for all $x \notin K$ (using the usual norm on F). Intuitively, there is a compact set outside of which $|f(x)|$ becomes vanishingly small. This is not quite the same as having compact support for f, but close. The set of continuous functions that vanish at infinity is denoted $C_0(X)$. It is easy to see that $C_C(X) \subset C_0(X) \subset C^0(X)$ as vector subspaces. As with $C^0(X)$, the relevant field F is implicit.

We'll use $C_C(M)$, so we'll need the following result:

Prop 6.1: The integral of any continuous compactly supported real function f on M is finite.

Pf: Let K be the compact support of f. The image of a compact set under a continuous map is compact, so $f(K)$ is compact in $\mathbb{R}$, and thus closed and bounded. Let $\alpha > 0$ be the bound on $|f(K)|$. Then $\int |f|\nu \leq \int \alpha\nu$. The volume of a compact region is finite, so $\alpha \int \nu$ is finite.

Note that this does *not* hold if f is just compactly supported. We need continuity. For example, $1/x$ on $[0,1]$ is compactly supported, but it has a discontinuity at 0 and is not in $L^1([0,1])$.

This tells us that a volume form ν on M is a linear functional on the vector space $C_C(M)$.

6.2. **Riesz representation theorem.**

To obtain a measure from a volume form, the key ingredient we need is the following:

Riesz-Markov-Kakutani representation Thm: Given Hausdorff, locally compact topological space X and positive linear functional $\eta : C_C(X) \to \mathbb{R}$, there exists a unique measure μ on some Σ containing the Borel algebra of X and s.t. (i) μ is finite on every compact subset of X, (ii) (Σ, μ) is measure-complete, (iiia) μ is outer regular, (iiib) μ is inner regular for open sets and sets of finite measure, and (iv) $\eta(f) = \int_X f\mu$ for all $f \in C_C(X)$.

See [13], p.40-47 (section 2.14) for a detail discussion and proof of this particular version of the theorem.

Let's decipher what this means, especially in the context of our problem.

- A "positive linear functional" on a vector space of functions is a linear functional η s.t. $\eta(f) \geq 0$ for any purely nonnegative function (i.e. if $f(x) \geq 0$ everywhere). $\hat{\mu}$ is obviously a positive linear functional whenever it is a linear functional. In particular, it is one on $C_C(M)$.
- As usual, the Hausdorff condition let's us speak of $\mu(K)$ for compact K. Without that requirement, property (i) would have no meaning since compact sets need not be measurable. We also require it for (iiib), since inner regularity involves a supremum over compact sets.
- Because X is Hausdorff and locally compact, property (i) is equivalent to μ being locally finite.
- Note the limitation on (iiib). In general, the Riesz theorem does *not* give us regularity (and thus a Radon measure). However, it does so in many circumstances. In our case, we will be seeking a finite measure. For any such measure, (iiib) is the same as ordinary inner regularity because the entire Borel algebra has finite measure. I.e. for a finite measure, (iiia-b) tell us that μ is regular.

  Put another way, if (iiib) doesn't translate into full inner regularity, then μ is not finite and our unique candidate for a satisfactory measure cannot be normalized into a probability measure. I.e., we will be out of luck. We'll have more to say about this below.

  This is *not* an iff. If μ is a finite measure, then (iiib) is true inner regularity. However, the converse need not hold. Suppose we have true inner regularity. This means that (iiib) is automatically satisfied since the condition holds for all measurable sets. However, this does not imply that all measurable sets are finite. It is quite possible that the inner-regularity condition also happens to be satisfied by sets of infinite measure. In fact, not even all the sets that fall under (iiib) need have finite measure; we could have open sets of infinite measure. We're not guaranteed that μ is finite.

  Note that local compactness does not guarantee that every open set has finite measure. It says that every point has an open neighborhood that sits in a compact set. Any such open set has finite measure, but such open sets by no means comprise the entire topology. There may be many open sets that do not sit in compact sets and do not have finite measure.
- Any manifold meets the topological criteria of being Hausdorff and locally compact.
- Any volume form on a differentiable manifold furnishes a positive linear functional on $C_C(M)$, as discussed.

- (ii) may seem superfluous. After all, there is always a unique completion of a given Borel measure. What really is being said is that we have a unique Borel measure with the specified properties. This then has a unique measure-completion (which produces some larger σ-algebra), and that completion also has the specified properties.

  If (i), (iiia-b), and (iv) hold for μ on the Borel algebra, they also hold for its completion. The integral is unchanged by the inclusion of sub-null sets. We saw in proposition 4.15 that, for a locally compact space, local finiteness and regularity both carry over to the completion (and it is easy to see that the same holds for (iiib) in its constrained form). We also saw in proposition 4.10 that finite measures complete to finite measures.

- The integral over X is just an ordinary Lebesgue integral. The Riesz theorem furnishes a σ-algebra Σ over X and a measure μ on Σ. Condition (iv) just involves the Lebesgue integral of f using Σ and μ.
- In our case, since η is our $\hat{\mu}$ built from the volume form ν, the integral condition really is just that $\int_M f\nu = \int_M f\mu$, with the integral on the left a differential-geometric integral and the integral on the right a measure-theoretic integral. I.e., μ just endows our Lebesgue integral with the same meaning as the differential-geometric one, at least for compactly-supported continuous functions on open subsets of M.

  In effect, μ extends the notion of volume embodied in ν beyond the topology to the entire Borel algebra and its completion.

- Note that it does not matter whether $\hat{\mu}$ is defined (or a linear functional) on a larger space of functions than $C_C(M)$. Since the theorem says "for every positive linear functional with property foo on space bar there is a unique blah measure", all we need do is to exhibit foo on space bar in order to get a unique associated blah measure. It does not matter whether foo can be defined on a larger space than bar.
- The unique (Σ, μ) restricts to a unique almost-Radon measure satisfying (iv) on $\mu|_{\Sigma_{Bor}}$.

  We don't mean "unique" in that no two μ's (ex. induced by distinct ν's on M) could restrict to the same measure, but rather that a given ν induces a specific almost-Radon measure that satisfies (iv) on Σ_{Bor}. The Riesz construction gives us (Σ, μ) as the completion of $(\Sigma_{Bor}, \mu|_{\Sigma_{Bor}})$ — so any other extension that isn't partway to the completion (by including only some sub-null sets) would require the addition of non-sub-null sets. I.e., we would need material changes to the σ-algebra beyond Σ_{Bor} and its sub-null sets.

We can summarize the theorem as saying that each volume form on M furnishes a unique "almost-Radon" measure on its Borel algebra. If that measure is finite, it is an actual Radon measure.

I.e., we'll use the term "almost-Radon" to describe a locally-finite measure that satisfies (iiia) and (iiib). An almost-Radon measure may fall short of true regularity — and thus fail to be a Radon measure — because of the limitation of (iiib) to open sets and sets of finite measure. For a finite measure, almost-Radon is synonymous with Radon.

Suppose we have a volume form ν on M and we take an open subset $O \subset M$. The following result tells us that it doesn't matter whether we restrict ν and then apply the Riesz theorem or apply the Riesz theorem and then restrict the resulting measure.

Prop 6.2: Given volume form ν on n-manifold M, and open subset $O \subset M$, let (i) μ be the almost-Radon measure on M obtained by the Riesz theorem from ν on M, (ii) let μ' be its restriction to O, and (iii) let μ'' be the almost-Radon measure on O obtained directly from $\nu|_O$. Then $\mu' = \mu''$.

Bear in mind that μ is on some σ-algebra Σ_M containing the Borel algebra of M s.t. (Σ_M, μ) is complete. Similarly, (Σ_O, μ') is complete for some Σ_O containing the Borel algebra of O.

Pf: (sketch) For brevity, we'll use "CSC" to mean continuous with compact-support. By proposition 2.6, O is an embedded n-submanifold of M. By proposition 4.8, their Borel algebras Σ_M^{Bor} and Σ_O^{Bor} obey $\Sigma_O^{Bor} \subseteq \Sigma_M^{Bor}$ (*not* a subalgebra on M, just a subset which is a σ-algebra on O). By proposition 4.14, the completion of the restriction is the restriction of the completion. As a result of all this, we need only show that $\mu' = \mu''$ on Σ_O^{Bor}. The Riesz theorem gives us an almost-Radon measure μ on Σ_M from ν on M. As a manifold, T is locally compact, so proposition 4.8 tells us that its restriction μ' to Σ_O^{Bor} is also an almost-Radon measure. [Technically, that proposition was proved for Radon measures, but it is trivial to see it holds for almost-Radon measures as well]. On the other hand, the Riesz theorem gives us an almost-Radon measure μ'' on Σ_O from $\nu|_O$. Let $i : O \to M$ be the inclusion map. Since i is continuous, it takes compact sets to compact sets. I.e., any compact set in O is compact in M. Because O is open in M, any compact subset of O has a buffer in the sense that it is contained in an open set of M. That means that any CSC function f on O extends to a CSC function f_e on M by setting its value to 0 outside O. $\int_M f_e \nu = \int_O f\nu$ because the support of f_e is limited to O, and $\int_M f_e \nu = \int_M f_e \mu$ via the Riesz theorem since f_e is CSC on M, and $\int_O f\nu = \int_O f\mu''$ via the Riesz theorem since f is CSC on O. Therefore $\int_M f_e \mu = \int_O f\mu''$. However, by proposition 4.8, $\mu' = \mu|_{\Sigma_O}$, so $\int_M f_e \mu = \int_O f\mu'$ since $f_e = 0$ outside O. We therefore have $\int_O f\mu' = \int_O f\mu''$. [Technically, this requires a bit more justification because we never proved that the integrals must be equal in such a case. However, it is fairly easy to see this from the definition of measure-theoretic integration — which we did not develop in these notes.] The Riesz representation theorem tells us that there is a unique almost-Radon measure on Σ_O that has this property for all CSC functions on O. I.e., $\mu' = \mu''$ because they agree on the integrals of all CSC functions on O.

As we'll see shortly, our manifold Ω_E has a natural volume form on it — so the Riesz representation theorem yields a unique measure satisfying (i-iv) on the Borel algebra of Ω_E. Okay, that's nice, but why do we want a unique measure that satisfies those particular constraints? Obviously, if we impose enough conditions on the measure we'll get something unique. We need to justify this particular choice. Let's now do so.

Put another way, suppose there are many measures induced by ν via the same "natural" mechanism. We may be able to prove that only one of them is ever a "blurble" measure, but that is irrelevant unless we specifically want a "blurble" measure. The mere fact of uniqueness isn't anything special. The Riesz theorem gives us existence and uniqueness. Existence of a measure satisfying (i-iv) is helpful because it tells us that we have at least one naturally-induced measure. However, uniqueness is only useful to us if we specifically require a measure with the properties (i-iv).

The need for (iv) is clear, since this is the key element that imbues $\hat{\mu}$ with our natural notion of volume. If we want our measure to reflect this, we need such a condition.

Our ultimate goal is a probability measure, and we thus require a finite measure. Conditions (i) and (iiia-b) stem from this, as we'll now show.

For this, we only need uniqueness up to a finite scale factor. Normalization maps all finite measures that differ by a scale factor to the same probability measure. If we have a natural choice of volume modulo a scale factor, or if our natural volume gives us a natural measure modulo a scale factor, we'll still end up with a unique probability measure.

Finiteness implies local finiteness, so any non-locally-finite measure is out of the running. Though local finiteness is not a sufficient condition for our purposes, it certainly is a necessary one.

As we mentioned already, if μ is finite, (iiib) is synonymous with true inner regularity. This means that for any finite μ, the Riesz theorem produces a Radon measure. By the contrapositive, if μ is not a Radon measure then it is not finite. I.e., from our perspective, requiring (iiia-b) is the same as requiring regularity.

But why require regularity at all? To anyone who has studied Lebesgue integration, it should have a familiar feel (in fact, the Lebesgue measure is often defined via a similar approach). However, that doesn't justify it per se. The following proposition helps.

Prop 6.3: Any finite measure on (the Borel algebra of) a locally compact, second-countable, Hausdorff space is regular.

See [14].

Suppose that our manifold is second-countable. Proposition 6.3 tells us that the *only* possible candidates for a finite measure are regular. Taken together with our previous comment on local-finiteness, we see that any finite measure must be locally finite *and* regular. I.e., it must be a Radon measure. Put another way, there are no finite non-Radon measures on the Borel algebra of a second-countable manifold. Since any Radon measure is almost-Radon as well, there are no finite non-almost-Radon measures either.

The Riesz representation theorem gives us a unique almost-Radon measure μ that adheres to the differential-geometric definition of integration (via condition (iv)). If Ω_E is second-countable and this μ fails to be finite we're out of luck; there can be no finite alternative, no matter how hard we look.

The converse does not hold. *If* we require second-countability, then we also require that μ be regular in order to have any hope (but not guarantee) of a finite measure. However, the need for a finite measure doesn't require this assumption. If we don't impose second-countability, we very well could have a finite measure that is not regular. Second-countability is an additional constraint. We must justify it, which we now will do. Note that the same argument does apply to (iiib) vs full regularity. In that case, if (iiib) doesn't amount to true inner-regularity, then μ cannot be finite.

Why should we require second-countability? We've just replaced the seemingly arbitrary condition of regularity with another seemingly arbitrary condition. However, second-countability is easier to justify as a requirement.

Proposition 2.5 tells us that for a Hausdorff manifold "second-countable" is equivalent to "paracompact with a countable number of connected components". By our definition of manifolds as paracompact and Hausdorff, this means that second-countable is equivalent to having a countable number of connected components. Q and Ω and the Ω_E's are all manifolds, so this holds for them.

Since the fiber $\mathbb{R}^n$ of T^*Q is connected, there is a trivial bijection between the connected components of Q and those of T^*Q. I.e., if Q has a countable number of them, so does T^*Q. Unfortunately, connectivity is not inherited in any simple way by $\Omega_E \subset T^*Q$.

If Q is connected, then T^*Q is connected, but it is quite possible that Ω_E is not. Nor can we rely on the continuity of H to help us; the inverse image of a connected set such as δ_E need not be connected, even under a continuous map.

However, proposition 2.2 tells us that second-countability *is* inherited. As a result, if Q has a countable number of connected components, Ω_E is second-countable. It then satisfies the conditions of proposition 6.3, and any finite measure on its Borel algebra must be regular.

Suppose Q has a countable number of connected components. We saw that T^*Q then has a countable number of connected components as well. By proposition 2.5, this is equivalent to T^*Q being second-countable. Proposition 2.2 then tells us that $\Omega_E \subset T^*Q$ is second-countable.

We'll either include second-countability of Q in our definition of a manifold or add a physics postulate (which is nonrestrictive in any meaningful physical sense) that the configuration manifold Q has a countable number of connected components.

As a practical matter, we usually take Q to be connected in classical mechanics. Even if not, it certainly is no great constraint to require a countable number of connected components.

Put simply, for any manifold Q with a countable number of connected components, any finite measure on the Borel algebra of Ω or Ω_E must be a Radon measure.

The Riesz representation theorem tells us that we get a unique Borel measure μ that is locally finite, outer regular, *almost* inner regular, and is consistent with the notion of integration embodied in the volume form ν. We are aiming for a finite measure, and any finite measure automatically is locally finite and (assuming second-countability of Q, which we now are) truly regular. I.e., any viable candidate which can be normalized to a probability measure must be a Radon measure. We are given a single almost-Radon candidate by the theorem and told that it is the *only* almost-Radon candidate. If it's finite, we are golden. If not, then it is impossible to construct a suitable probability measure. No matter how clever we try to be, and no matter how much domain-specific knowledge we employ, we cannot find an alternative because none exists.

The Riesz theorem gives us a unique locally finite, almost-Radon (Σ, μ) that satisfies (iv), is complete, and where Σ contains Σ_{Bor}. Strictly speaking, much of what we said above applies to measures on Σ_{Bor}. However, the completion only adds sub-null sets with measure zero and doesn't change the measures of any other sets. I.e., if (Σ, μ) is the completion of $(\Sigma_{Bor}, \mu|_{\Sigma_{Bor}})$ then all the relevant properties are inherited — and not just inherited but with an iff. μ is FOO on σ iff $\mu|_{\Sigma_{Bor}}$ is FOO on Σ_{Bor}, where FOO is "finite", "regular", "almost-regular", "Radon", or "almost-Radon". We have no extra flexibility, because Σ_{Bor} is smaller than Σ only in ways that don't matter. As an example, proposition 6.3 may speak of measures on Σ_{Bor} but it applies equally well to their completions.

It turns out that the physical assumptions we have imposed so far are not sufficient to guarantee the finiteness of μ. We are forced to adopt an additional constraint, albeit a very reasonable and mild one. We will do so shortly, but let's first understand how our volume form arises.

At this point, we may wonder whether a volume form is too restrictive. After all, a volume form must be positive, but a measure need only be nonnegative. All volumes defined by ν on a manifold must be positive, but a measure can have measure-zero sets. Shouldn't the corresponding notion be a nonnegative n-form? There are two issues at play here. First, we never said that every finite measure on the Borel algebra defines a volume form (it does, but we never said so). The more relevant consideration is *which* sets the volume form defines a volume on. It only defines the volumes of open sets, but proposition 2.6 tells us that every nonempty open set is an n-dimensional submanifold of M. As a result, there are no measure-zero nonempty open sets in the Borel algebra (or in its completion). The open subsets of a manifold have heft.

7. Review of symplectic manifolds

Let's now see how a natural volume form arises on Ω, and thus on Ω_E. Our Ω is a cotangent bundle, and every cotangent bundle has a symplectic structure via the Poincare 2-form (which we'll discuss shortly). We'll therefore make a brief digression to review symplectic manifolds.

We'll begin with the definition of a symplectic form on a vector space. This is most easily understood in contrast with the more familiar notion of an inner product, so we review that and draw the comparison. Then we define a symplectic form on a manifold and discuss the tautological one and two forms on a cotangent bundle.

A symplectic form ω on finite-dimensional vector space V over field K is a map $\omega : V \times V \to K$ that is:

- (i) Bilinear.
- (ii) Nondegenerate: $\omega(v, w) = 0$ for all w iff $v = 0$.
- (iii) $\omega(v, v) = 0$ for all v.

Property (iii) is equivalent to anti-symmetry ($\omega(v,w) = -\omega(w,v)$) for fields of characteristic $\neq 2$. In particular, the two are equivalent for $K = \mathbb{R}, \mathbb{C}$ (which both have characteristic 0). We'll exclusively consider real symplectic forms (i.e. $K = \mathbb{R}$) here.

One can define a notion of symplectic form on an infinite-dimensional space, but it is far more complicated. We'll assume a finite dimension in these notes.

Prop 7.1: A symplectic form can only be defined on an even-dimensional vector space.

This will be most evident when we discuss Darboux coordinates.

Pf: Pick some $v_1 \neq 0$. Since $v_1 \neq 0$ and ω is non-degenerate, there exists some $v_1' \neq 0$ s.t. $\omega(v_1, v_1') \neq 0$. Since $\omega(v,v) = 0$ and ω is bilinear, we know that v_1' is not proportional to v_1. I.e., they are linearly independent. We thus have a pair of basis elements for V. Next, consider the subspace $V' \subset V$ consisting of all $v \in V$ that are linearly independent of v_1 and v_1'. A vector is linearly dependent on v_1 and v_1' iff it satisfies $\omega(v, av_1 + bv_1') = 0$ for some a and b. I.e., we may write $V = V_1 \oplus V'$, where $V_1 = span(v_1, v_1')$ and V' has all vectors linearly independent of them. V' therefore consists of all $v' \in V$ s.t. $\omega(v', V_1) = 0$ (meaning that $\omega(v', v) = 0$ for every $v \in V_1$). We next repeat this process on V' to obtain $V' = V_2 \oplus V''$, and so on. Since V is finite-dimensional, we eventually run out of pairs. If V is odd-dimensional, we would end up with a final one-dimensional vector space — call it V^f. Suppose $v' \in V^f$ is nonzero. By construction, $\omega(v', v) = 0$ for all $v \notin V^f$, and by (iii) $\omega(v', v) = 0$ for all $v \in V^f$. I.e. $\omega(v', v) = 0$ for all $v \in V$ but $v' \neq 0$, which violates (ii). Therefore, V must be even-dimensional.

7.1. **Brief review of Inner Products.**

A symplectic form can be thought of as the antisymmetric counterpart of an inner product, so let's briefly recall some salient features of inner products. An inner product α on a real vector-space V is a map $\alpha : V \times V \to \mathbb{R}$ that is (i) bilinear, (2) symmetric, and (3) positive definite.

In fact, real inner products and symplectic forms are closely related in another way. Recall that a Hermitian inner product on a complex vector space V is a map $H : V \times V \to \mathbb{C}$ that is (i) linear in the first argument ($H(cv + w, u) = c \cdot H(v,u) + H(w,u)$ for $c \in \mathbb{C}$), (ii) conjugate symmetric ($H(v,w) = \overline{H(w,v)}$), and (iii) positive definite ($H(v,v) > 0$). The first two imply that it is conjugate linear in the 2nd argument: $H(v, cw + u) = \overline{c} \cdot H(v,w) + H(v,u)$. Given a Hermitian inner product H, we may write it as $H = \alpha + i\omega$. If we view V as a $2n$-dimensional real vector space in the usual way, then α is a (real) inner product on it and ω is a (real) symplectic form on it.

Any inner product α on a real vector space V defines the following:

- A special class of bases called the "orthogonal" bases, for which the basis vectors satisfy $\alpha(b_i, b_j) = c_i \delta_{ij}$ for some $c_i > 0$.
- A special class of linear automorphisms on V called "isometries" (aka "orthogonal transformations"), which preserve the inner product.

  I.e. an isometry is an invertible linear map $T : V \to V$ s.t. $\alpha(T(v), T(w)) = \alpha(v,w)$.
- A special class of bases called the "orthonormal bases", for which the basis vectors satisfy $\alpha(b_i, b_j) = \delta_{ij}$.

  Obviously, these are orthogonal bases as well.

The general linear automorphisms on V form the group $GL(V)$, and the isometries form the subgroup $O(V)$. Each linear automorphism can be viewed as either a "passive" change of basis or an "active"

transformation of vectors.

In linear algebra, the terminology can be problematic. Points in V are referred to as vectors and points in $GL(V)$ are referred to as linear operators (or linear transformations), and both are independent of basis. In a given basis, the numeric expansion of a linear operator in terms of the basis elements is a matrix — so there is no ambiguity in the language. However, the numeric expansion of a vector is also referred to as a vector. This can make it difficult to linguistically distinguish a vector as an abstract object from an n-tuple of coefficients in a particular basis. We'll therefore use the term matrix for numeric arrays, whether a column or row or square, and reserve the term "vector" for the abstract object. A $1 \times n$ matrix is a row, an $n \times 1$ matrix is a column, and an $n \times n$ matrix is a square.

Under a basis change, we replace our choice of basis vectors $B = (b_1, \cdots, b_n)$ with some other $B' = (b'_1, \ldots, b'_n)$, where B and B' each is a set of linearly independent vectors. A vector $v \in V$ doesn't change, but its coefficient expansion in terms of the new basis is different. On the other hand, under a transformation our choice of basis vectors remains the same but v actually changes. Its coefficients in our basis B therefore change as well.

Consider basis $B = \{b_1, \ldots, b_i\}$ and linear automorphism $T : V \to V$ and vector $v \in V$. In basis B, v has a unique expansion $v = \sum c_i b_i$ for some coefficients c_i in the relevant field K (which we'll take to be $\mathbb{R}$ for the current discussion). I.e., $c^T \equiv (c_1, \ldots, c_n)$ is the row matrix for v in the basis B, and c is its column version. Tv is a new vector, whose expression in basis B is $Tv = \sum c'_i b_i$ for some c'. If the expression for T in basis B is the square matrix M, then $c' = Mc$, a matrix equation. This is the active transformation.

Under a change of basis, the row (and column) matrix for each vector changes *and* the square matrix for every transformation changes. Suppose we express B in terms of B' as $b_i \equiv S_{ji} b'_j$. Then S is a square matrix whose i^{th} column is just the column-matrix for vector b_i in the basis B'. If c is the column matrix for v in B, then $c' = Sc$ is the column matrix for v in B'. Suppose some linear operator T has (square) matrix expression M in basis B. Then the matrix expression for Tv is Mc in basis B and SMc in basis B'. If M' is the matrix expression for T in B' then Tv has expression $M'Sc$ in basis B'. These two must be equal, so $SMc = M'Sc$ for all c, which gives us $M' = SMS^{-1}$. This is the basis-change behavior of the matrix for a linear operator T.

Given any $T \in GL(V)$, we may regard T as either a basis change or a transformation. Suppose its expression in basis B is the (square) matrix M. As an active transformation, it takes each b_i to some b'_i (as it does for any vector v). Since the column matrix for b_i in basis B is just $c^j_i = \delta^j_i$ (where we use a superscript for clarity only), the expression for b'_i in basis B is $c'^j_i = M_{jk} c^k_i = M_{ji}$. I.e., $b'_i = M_{ji} b_j$. This tells us that M acts as the inverse of the corresponding S. This makes sense, since S expresses B in terms of B'. Put another way, a linear operator viewed as an active transformation has the inverse matrix expression as when viewed as a basis change. This isn't because its coefficients differ in a given basis, but rather how we use them. In one case, we use M and in the other we use $S \equiv M^{-1}$.

We can use the isometry group $O(V)$ to partition the set of all bases into classes via the equivalence relation: $B \sim B'$ iff the basis change is an isometry. The orthogonal bases form a single equivalence class.

In any basis, if $T \in O(V)$, its matrix representation M (and thus the corresponding basis-change representation $S = M^{-1}$) is an orthogonal matrix — meaning that $M^T M = I$. I.e. $M^{-1} = M^T$.

In any orthogonal basis, the matrix for the inner product α is diagonal. Its diagonal elements are the lengths of the basis vectors. What do we mean by this? An inner product induces a norm on V, and the length of a vector is $\sqrt{\alpha(v,v)}$.

Since isometries don't change $\alpha(v,w)$, they don't change the norms of vectors. The lengths of vectors therefore don't change under an isometry-induced basis change, and this includes the lengths of the basis vectors themselves. All that changes is our choice of basis vectors — and the new basis vectors can have different lengths from the old ones. But wait — the eigenvalues of a matrix are invariant under basis changes, and the eigenvalues of a diagonal matrix are just its diagonal elements. At most, a basis change should permute them. What gives? The problem is that we are using "matrices" to represent two distinct types of objects, and these transform differently under a basis change. As we saw, the matrix representation of a linear operator transforms as $M' = SMS^{-1}$. Consider the inner product α, and suppose its matrix representation is M in B and M' in B'. In order for $\alpha(v,w)$ to be invariant under basis change S, we need $c'^T M' d' = c^T M d$, where c and c' are column vector representations of v in B and B' and d and d' are those for w. The primed expression becomes $c^T S^T M' S d = c^T M d$. Since this must hold for all c and d, we have $M' = (S^T)^{-1} M S^{-1}$. This is a different transformation rule than that of a linear operator. How can a matrix transform in two distinct ways? We've been very sloppy in our use of the term. As a numeric array, a "matrix" doesn't "transform" at all. It's just an array of numbers. If we want to speak of matrices as "transforming" under a basis change, we must provide a rule that maps matrices to matrices. This rule is external information and depends on the type of object the matrix represents. What we think of as "ordinarily behavior" is that of the matrix representation of a linear operator, aka a $(1,1)$-tensor. On the other hand, a bilinear form such as the inner product is a $(0,2)$-tensor (meaning that it takes two vectors as arguments and returns a scalar). Both α and T have square-array representations in a given basis, and we can use matrix algebra in the ordinary way to work with both — but we must be very careful to distinguish their behaviors under basis changes. They are fundamentally different types of objects. This distinction is obscured when confining ourselves to isometries — since $(S^T)^{-1} M S^{-1} = SMS^{-1}$ in that case — leading to the impression that this is universal. It is not, and their behaviors differ in general. Returning to our original concern, we can define eigenvalues for any numerical matrix. However, these are only invariant under similarity transforms of matrices. The matrix for a linear operator changes via a similarity transform under a basis change, but the matrix for an inner product does not. Therefore, the eigenvalues of the matrix representation of an inner product are not meaningful because they are basis-dependent.

Within the class of orthogonal bases, there is a subclass of orthonormal bases. In an orthonormal basis, the inner product takes the form of the identity matrix.

Note that there is no such thing as an "orthonormal matrix" or an "orthonormal linear operator" or an "orthonormal subgroup of isometries". The only orthonormal concept is that of a basis. Similarly, the notion of orthonormality does *not* lead to a refinement of the partition (on the set of all bases) into smaller classes of bases. The orthonormal bases form a subset of the class of orthogonal bases, nothing more.

7.2. **Symplectic forms on Vector Spaces.**

Now let us consider a symplectic form ω on V. Just as an inner product gives rise to a notion of orthonormal bases and orthogonal transformations, a symplectic form gives rise to a notion of symplectic bases and symplectomorphisms (aka canonical bases and canonical transforms). A symplectomorphism is an isometry of the symplectic form. I.e., it is an automorphism that preserves the symplectic form.

More generally, let (V,ω) and (V',ω') be two vector spaces equipped with symplectic forms. If $T : V \to W$ is an invertible linear map, we require $\omega'(Tv,Tw) = \omega(v,w)$ for all vectors v, w. We'll be focusing on automorphisms, so $V' = V$ and $\omega' = \omega$. However, when we discuss symplectic manifolds, we'll need to consider more general symplectomorphisms. Note that we could have generalized in a similar manner our discussion of inner products, speaking of invertible linear maps and inner-product preserving isometries between different inner-product spaces. We chose to provide the bare minimum necessary for our purposes, so we stuck with "transformations" (and basis changes), which are automorphisms.

Just as an orthonormal basis $(e_1, \ldots, e_n)$ is one in which $\alpha(e_i, e_j) = \delta_{ij}$, a symplectic basis $(e_1, \ldots, e_n, f_1, \ldots, f_n)$ is one in which $\omega(e_i, e_j) = 0$, $\omega(f_i, f_j) = 0$, and $\omega(e_i, f_j) = \delta_{ij}$. I.e., any symplectic basis has an intrinsic pairing of basis vectors (e_i, f_i).

There are two ordering conventions: (i) $(e_1, f_1, e_2, f_2, \dots)$ and (ii) $(e_1, \dots, e_n, f_1, \dots, f_n)$. Each is useful for some purposes and clumsy for others. Most of the time, (ii) is preferable because it allows transformation matrices to be represented as 4 big $n \times n$ blocks rather than n small 2×2 blocks. Under convention (ii), in any symplectic basis, ω takes the matrix form $J \equiv \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, where I is the $n \times n$ identity matrix. I.e., $\omega(v, w) = v^T J w$ in such a basis.

The counterpart of "orthogonal" bases (which presumably would involve $\omega(e_i, f_i) = c_i$ for some nonzero $c_i > 0$) doesn't come up, so it apparently is not useful.

7.3. **Symplectic manifolds.**

A "symplectic form" on a manifold M is just a closed nondegenerate 2-form. A manifold with a symplectic form is termed a "symplectic manifold".

Recall that a nondegenerate 2-form has $\omega(v, w) = 0$ for all w iff $v = 0$, and a closed form ω has $d\omega = 0$.

The terminology is not ideal. One speaks of an "inner product" on a vector space and a "Riemannian metric" on a manifold, so there is no room for confusion. However, the term "symplectic form" is used for both the object on a vector space and the object on a manifold.

Like a Riemannian metric, a symplectic form on a manifold comprises additional structure. Even if there is a unique or natural symplectic form that we obviously would use, we still must be clear that we are working in the category of symplectic manifolds rather than that of manifolds. We'll write (M, ω) for a symplectic manifold.

It is easy to see that a symplectic form on M is just a suitably differentiable choice of (vector space) symplectic form on each tangent space T_pM. Consider $\omega_p : T_pM \times T_pM \to \mathbb{R}$. As a 2-form, it is bilinear and antisymmetric. Nondegeneracy is a point-wise property, so this carries over to ω_p as well. Therefore, ω_p is a symplectic form on T_pM. Note that the converse is not true. If we have a symplectic form on each T_pM, these don't necessarily combine into a symplectic form on M. This is true even if they vary smoothly. A symplectic form must also be closed, and this involves a specific set of differential equations for the coefficient functions in each chart.

Following up on the previous comments, why *do* we require ω to be closed? It turns out that if we omit this condition, we get "almost-symplectic" forms, which are quite useful for many purposes. *If* $d\omega = 0$, we get a lot of nice properties (ex. Darboux's theorem) and consequences that align with the laws of physics. I.e., this is the mathematical object that best codifies certain physics. That isn't necessarily a justification for *demanding* closedness in the definition, but since the objects of interest to us stem from their utility, defining it this way isn't unreasonable. In physics, the relevant symplectic form *is* closed. As we will see in a moment, it is the Poincare 2-form on T^*Q — which is not only closed, but exact. It would be natural to adopt this as the template for our definition. Requiring exactness turns out to be unnecessary, and we get everything we need with closedness — so this is probably why the definition includes that. Ultimately, it is just a matter of convention and nomenclature.

A "symplectomorphism" is a diffeomorphism which preserves the symplectic form. If a symplectomorphism exists between (M, ω) and (M', ω'), they are considered the same from a symplectic standpoint.

Formally, a diffeormorphism $f : M \to M'$ is a symplectomorphism if $f^*\omega' = \omega$, with f^* the pullback.

A given manifold may have no symplectic forms or many symplectic forms. If it has symplectic form ω, then $c\omega$ is symplectic for every $c \neq 0$. Note that (M, ω) and $(M, c\omega)$ are *not* related by a symplectomorphism in general.

It is easy to see that a symplectomorphism f induces a (vector space) symplectomorphism $T : T_pM \to T_{f(p)}M'$. As a diffeomorphism, f defines an isomorphism $f_*|_p : T_pM \to T_{f(p)}M$ via $(f_*|_p v)(g) \equiv v(g \circ f)$ for $v \in T_{f(p)}M$ and any smooth fn g on M'. Since $f^*\omega' = \omega$, and f^* is defined as $(f^*\omega')(v, w) \equiv \omega'(f_*v, f_*w)$, we see that $\omega'(f_*v, f_*w) = \omega(v, w)$. Our invertable linear map T is $f_*|_p : T_pM \to T_{f(p)}M'$, and it does indeed satisfy $\omega'_{f(p)}(Tv, Tw) = \omega_p(v, w)$.

Prop 7.2: Darboux's Theorem: Given a symplectic $2n$-manifold (M, ω), every point $x \in M$ is contained in some chart where (in the coordinate basis) $\omega = \sum dq^i \wedge dp^i$.

See [4], corollary 8.1.3 for a discussion and proof.

What we mean by this is that we can choose a pairing of the coordinates in such a chart so that the expression is as stated.

These are called "Darboux coordinates" or "Darboux charts". It is easy to see that, in them, each ω_p has the canonical matrix form J on vector space T_pM when expressed in the corresponding coordinate basis. I.e., the coordinate basis for Darboux coordinates is a symplectic basis.

This m
choice o

In the context of classical mechanics, "canonical coordinates" are just Darboux coordinates associated with the natural symplectic form on T^*Q (which we'll describe shortly).

Bear in mind that ω is a symplectic form regardless of its expression in local coordinates. In Darboux coordinates (i.e. in those charts of the maximal atlas that are Darboux charts), it happens to take the nice succinct form we are used to. This makes these attractive to work in.

Suppose we have two symplectic forms ω and ω' on M, each with its own associated Darboux atlas (with both atlases subsets of the maximal atlas that forms the differential structure on M). Even on a smooth real manifold, it is possible to have $\omega = \omega'$ on an open set but $\omega \neq \omega'$ elsewhere. I.e., the atlases need not be disjoint and can have Darboux charts in common. They can even have the same expressions in those charts.

Just as we lose no generality by sticking to a smooth atlas, we lose none by sticking to a Darboux atlas. This is why some treatments (misleadingly) act like Darboux charts are the only charts or symplectic forms always have this simple expression.

Prop 7.3: The converse holds too. Given a 2-form ω on M, if there exists a Darboux atlas for it (i.e. if we can cover M with Darboux charts of ω), then ω is a symplectic form on M.

Pf: Given any $x \in M$, by assumption there is some Darboux chart containing it. In this chart, $\omega = \sum dq^i \wedge dp^i$, where we've ordered the coordinates as needed. This trivially is closed and nondegenerate on the chart. Since ω (as a 2-form) is independent of the chart, this means it's closed and nondegenerate on an open set containing x. This is true for all x, so we have our result.

Any symplectic manifold (M, ω) has a natural volume form ω^n. This means that M is orientable and has a natural orientation which makes ω^n positive on every connected component.

This is easy to see in Darboux coordinates. If we pick a particular Darboux chart as our starting point, then there is a unique maximal oriented atlas containing it. There also is a unique maximal Darboux atlas containing it. The intersection of these is an oriented Darboux atlas (i.e. it covers Ω). Note that if we chose a different starting Darboux chart, we could end up with a different orientation.

This canonical volume actually is only specified up to a factor. Some people use ω^n, but many define the volume form to be $\frac{\omega^n}{n!}$. The factor amounts to a choice of units. For our purposes, it is irrelevant because we're deriving a probability measure. Any scale factor will be washed away when we normalize our volume-derived measure.

If we replace ω with $c\omega$ (for some $c \neq 0$), then the volume form would become $c^n\omega^n$. Note that it doesn't matter if the sign is $-$ (as is the case when $c < 0$ and n is odd), because the sign of ω^n is arbitrary anyway. In fact, the most general form would be $a \cdot c^n \cdot \omega^n$, where $a \neq 0$ is the scale factor we imposed in our previous comment, and c is the scale from changing the symplectic form. If we pick a sign, there exists an orientation that makes $a \cdot c^n \cdot \omega^n$ a volume form. As with a, any choice of c will be washed away when we normalize our volume-derived measure into a probability measure.

Prop 7.4: Any open subset of a symplectic manifold is a symplectic manifold.

Pf: An open subset of M is an embedded submanifold of the same dimension as M, so we can just restrict ω to it. Nondegeneracy and closedness are preserved, since both are local properties.

A quick note on complex manifolds: It may be tempting to imagine that any complex manifold is symplectic when viewed as a $2n$-dimensional real manifold, using the obvious split into real and complex components as its symplectic pairing. This is not the case.

$\mathbb{C}^n$ is symplectic when endowed with the usual Hermitian inner product, but not as a plain old differentiable manifold. More generally, a complex manifold by itself is not symplectic, but a complex Hermitian manifold (i.e. a complex-analytic manifold with a specific Hermitian metric on it) is symplectic when viewed as a $2n$-dimensional real manifold in the usual manner. In that case, the symplectic form is just the imaginary part of the Hermitian metric, taken point by point as described earlier.

7.4. T^*Q as a symplectic manifold.

Given any differentiable manifold Q, its cotangent bundle T^*Q is a symplectic manifold with a natural symplectic form. We will need this form, so let's see how it arises.

Intuitively, the basic idea is simple: we have a natural pairing of coordinates on T^*Q given by (with lots of caveats to be discussed below) associating a local coordinate q_i on Q with the corresponding basis vector dq_i in the vector space T^*_qQ.

7.5. **Tautological 1-form on T^*Q.**

Any cotangent bundle T^*Q has a special 1-form (variously called the "tautological" or "canonical" or "fundamental" or "Poincare" 1-form).

As a vector bundle, T^*Q has a canonical projection $\pi : T^*Q \to Q$ that takes each cotangent space T^*_qQ to the underlying point $q \in Q$ via $\pi(q, \omega_q) = q$. This induces a push-forward $\pi_* : T(T^*Q) \to TQ$ and a pull-back $\pi^* : T^*Q \to T^*(T^*Q)$.

At each point $(q, \omega_q) \in T^*Q$, π_* takes $T_{(q,\omega_q)}(T^*Q)$ to T_qQ. As such, it is a point-by-point map between tangent spaces. It does *not* take vector fields to vector fields. The map π is non-injective, so pushing forward a vector field would require us to select which point's tangent space we map to T_qQ. The pull-back π^* of a 1-form suffers no such deficiency because a function can't be one-to-many, so we can pull back 1-forms as fields.

Formally, the push-forward is given by $(\pi_* v)(f) = v(f \circ \pi)$ and the pull-back is given by $\pi^*\omega(v) = \omega(\pi_* v)$.

Consider a given point $z = (q, \omega_q)$ in T^*Q. ω_q is a linear map $T_qQ \to \mathbb{R}$. Define covector θ_z via $\theta_{(q,\omega_q)}(v) \equiv \omega_q(\pi_* v)$ (where $v \in T_{(q,\omega_q)}(T^*Q)$). I.e., θ_z is the pull-back of ω_q at that point.

This defines a special 1-form on T^*Q, obtained by pulling back the 1-form on Q which each point in T^*Q represents. Even though constructed point-by-point in this fashion, the resulting object is indeed a 1-form on T^*Q.

I.e. θ, expressed in each chart's coordinates, is suitably differentiable for the manifold. It patently looks like a 1-form at each point, so this is the only potential obstruction to being a 1-form on T^*Q.

Note that θ isn't the pull-back of a 1-form on Q. Rather, we have pulled back each cotangent space in a way which relies on the ω_q for the underlying point in T^*Q. Why do this? A section of T^*Q is a 1-form on Q. If ω is such a 1-form on Q, then $\pi^*\omega$ is a 1-form on T^*Q, aka a section of $T^*(T^*Q)$. It replicates ω_q on each point in the fiber over q, effectively discarding n of the degrees of freedom we have. However, only one point in that fiber over q actually corresponds to (q, ω_q). Any given 1-form on Q yields this sort of uninteresting 1-form on T^*Q. We don't have a preferred 1-form on Q, but we do have the extra n degrees of freedom to compensate for that. By using these, we can build a special 1-form on T^*Q point-by-point in the fashion described.

θ is the unique 1-form on T^*Q which is consistent with the "meaning" of T^*Q as a manifold of 1-forms. I.e., for every (q, ω_q), $\theta_{(q,\omega_q)}$ is the pull-back of ω_q along π.

Let's see what's really going on here at the coordinate level. A 1-form θ_z at the point $z \equiv (q, \omega_q) \in T^*Q$ is a linear function on the tangent space of T^*Q at z. If we pick a basis $\{e_i\}$ for T^*_qQ, then $\omega_q = \sum c_i e_i$ for some coefficients c_i. In a given chart (U, ψ) on Q, we have a set of coordinates $(q_1 \dots q_n)$. These induce coordinate bases $(\partial_1, \dots, \partial_n)$ for T_qQ and $(dq^1, \dots, dq^n)$ for T^*_qQ.

A tangent vector is a linear map from smooth functions on Q to $\mathbb{R}$, and $\partial_i f = \frac{\partial f}{\partial q_i}$ at each point $q \in U$, with ∂ on the right side denoting the ordinary partal derivative in $\mathbb{R}^n$. Similarly, dq^i is a linear map from $T_qQ \to \mathbb{R}$, given by $dq^i(\partial_j) = \delta^i_j$. I.e., if we pick a specific set of coordinates and use the corresponding coordinate basis for T^*_qQ, then $\omega_q = \sum c_i dq^i$ for some coefficients c.

Denote by $V^* = \mathbb{R}^n$ the canonical fiber for T^*Q. I.e. T^*_qQ is diffeomorphic (as a manifold) and isomorphic (as a vector space) to V^* for every $q \in Q$. Because T^*Q is a vector bundle, it locally looks like $O \times V^*$ for

some $O \subseteq Q$. Around each point $q \in U$, there is some such O, so we'll confine ourselves to a chart that restricts ψ to $O' \equiv O \cap U$. Within O', $\pi^{-1}(O')$ looks like $O' \times \mathbb{R}^n$, with $(dq^1, \ldots, dq^n)$ the common basis for all the fibers. Any vector in V^* can be written $\omega = \sum c_i dq^i$ for some set of real c_i's. These c_i's are the "coordinates" for V^*. I.e., our local coordinates for T^*Q are just $(q_1, \ldots, q_n, c_1, \ldots, c_n)$.

Note that the dq^i are *not* coordinates. They are basis vectors. The coordinates which parametrize V^* are the coefficients when we expand in those basis vectors.

Consider $T_z(T^*Q)$, the tangent space to T^*Q at $z = (q, \omega_q)$. Since T^*Q locally looks like $O' \times V^*$, $T_z(T^*Q)$ separates into a product of tangent spaces. Any vector at z can be written (v, w), with $v \in T_qQ$ just a tangent vector to Q at q, and w a tangent vector to the copy of V^* over q.

A 1-form at z takes $T_z(T^*Q)$ to $\mathbb{R}$. However, we have a natural map of this sort since the point z contains the 1-form ω_q. As a 1-form at q on Q, ω_q can eat the v part of our (v, w) tangent vector. Like the tangent space at z, the cotangent space $T_z^*(T^*Q)$ also is a product space. This means that we can define a 1-form $(\omega_q, 0)$ which takes (v, w) to $\omega_q(v)$. I.e., we just use the 1-form part of z to define a 1-form on T^*Q by ignoring the tangent to the fiber. The result is the canonical 1-form θ.

Since the local coordinates we've chosen on T^*Q are $(q_1, \ldots, q_n, c_1, \ldots, c_n)$, the corresponding coordinate basis for $T_z^*(T^*Q)$ would be written $(dq^1, \ldots, dq^n, dc^1, \ldots, dc^n)$. Any 1-form on T^*Q at z can be expressed $\omega_z = \sum a_i dq^i + \sum b_i dc^i$.

For our special 1-form θ, the b_i's are all zero, and $\theta_z = \sum a_i dq^i$. However, these a's are just those of ω_q (from $z = (q, \omega_q)$) when expressed in the coordinate basis on T_q^*Q. I.e., if $\omega_q = \sum c_i dq^i$, then $\theta_z = \sum c_i dq^i$.

This notation is a little deceptive, because dq^i is doing double duty as a basis vector in two distinct vector spaces. If we write $T^*_{(q,\omega_q)}(T^*Q) = W_1 \times W_2$, where W_1 is the part acting on T_qQ and W_2 is the part acting on vectors tangent to the fiber, then there is a natural isomorphism between W_1 and T_q^*Q. Formally, this is obtained via pull-backs and push-forwards along the projection map. The natural isomorphism takes dq^i in one vector space to the corresponding dq^i in the other.

The c's parametrize the cotangent fiber over q and are commonly denoted $(p_1, \ldots, p_n)$. To avoid allowing preconceptions to creep in, I chose to call them something different so far. We'll now revert to the ordinary usage. Replacing the notation c_i with the usual p_i, we have $\theta = \sum_i p_i dq^i$.

The interpretation of this expression requires care. The dq^i here are one-forms on T^*Q. If $(q_1, \ldots, q_n, p_1, \ldots, p_n)$ are our coordinates for $\pi^{-1}(O) \subset T^*Q$, then the dq^i are local one-forms. As such, they can serve as basis vectors for the cotangent spaces — much as we did before. The p's are coordinates on an equal footing with the q's, and we can define the local coordinate one-forms dp^i. Each p_i is (locally) a function on T^*Q, just as each q_i is. It is *not* a function on Q. I.e., dp^i is not the differential of a function on Q, and it cannot be expressed in terms of the dq^i's as if it were.

Prop 7.5: θ is not closed. I.e. $d\theta \neq 0$.

Pf: As mentioned, the p's and q's are on an equal footing as coordinates for T^*Q. Each dp^i is an independent one-form, linearly independent everywhere from the dq^i's. As such, $d\theta = \sum_i dp^i \wedge dq^i$. Once again, p_i is not a function on Q and we cannot express its derivative in terms of the dq^i's. The expression $d\theta$ is patently nonzero.

7.6. **Tautological 2-form on T^*Q.**

Since the tautological 1-form θ on T^*Q is not closed, its exterior derivative is not zero. We can take this derivative to get a tautological 2-form $\omega \equiv -d\theta$. As we saw, in (p, q) coordinates, $d\theta = \sum dp^i \wedge dq^i$. The

minus sign is just a convention to put the q's before the p's, as is common in physics. We therefore have $\omega = \sum dq^i \wedge dp^i$.

As with θ, ω is variously referred to as the "tautological" or "canonical" or "fundamental" or "Poincare" 2-form. It is a natural 2-form on the cotangent bundle of every differentiable manifold, and it has some very nice properties.

ω looks symplectic, but does that mean it has the Darboux form in every possible chart? If so, the symplectic atlas would be the entire maximal manifold atlas for T^*Q. The answer is no. The Darboux form is an artifact of how we constructed the p's from the q's.

Given any chart (U, ψ) of Q and any open set $O \subset U$ s.t. T^*Q looks like a product space over O, there is a corresponding manifold chart for T^*Q in which ω has this canonical form. However, such charts form a tiny subset of the total (manifold) charts of T^*Q.

The chart is defined using $(q_1, \ldots, q_n, p_1, \ldots, p_n)$ as its coordinates. Let (U, ψ) embody this chart (with $U \subset Q$ and $\psi : U \to \mathbb{R}^n$ the homeomorphism to a subset of $\mathbb{R}^n$). From this, we can obtain both a manifold and vector-bundle chart for T^*Q over $O' \equiv U \cap O$. The manifold chart is $(\pi^{-1}(O'), \psi')$, with $\psi'(q, \omega_q) \equiv (q_1, \ldots, q_n, p_1, \ldots, p_n)$, where $(q_1, \ldots, q_n)$ is just $\psi(q)$ (i.e. the coordinates conferred by ψ) and $p_1, \ldots, p_n$ are the coefficients (in the coordinate basis) of ω_q. I.e., $\omega_q = \sum p_i dq^i$. The vector-bundle chart is $(\pi^{-1}(O'), \phi)$, with $\phi(q, \omega_q) = (q, (p_1, \ldots, p_n))$.

Not all manifold charts of T^*Q arise this way. Some may involve open sets that contain pieces of fibers (i.e. which aren't $\pi^{-1}(U)$ for some open U in Q) or which involve nonlinear parametrizations of the covector spaces over points. Such charts obviously wouldn't also serve as vector bundle charts since they lose either the product structure or the linear parametrization of the fiber (which is part of the definition of a vector bundle atlas). Suppose we just consider charts of the form $(q_1, \ldots, q_n, c_1, \ldots, c_n)$, obtained by picking some non-coordinate basis $v^1, \ldots, v^n$ for V^*, expanding each $\omega_q = \sum c_i v^i$, and using the coefficient c's as our parametrization. Such charts can serve in both the manifold and vector-bundle atlases for T^*Q because they preserve the local product structure and are linear parametrizations. However, ω won't take the canonical form $\sum dq^i \wedge dp^i$ in them. Nor does $\theta = \sum p_i dq^i$ in these charts.

Since we have a set of (manifold) charts which cover T^*Q and in all of which ω takes the Darboux form, T^*Q is a symplectic manifold.

We could also show this directly. ω is clearly bilinear and antisymmetric. As an exact differential form, it also is closed. Showing that it is non-degenerate is not difficult either.

8. Putting it all together

We're now in a position to assemble all this machinery into a notion of uniform probabities, usable by the AEAP. Let's briefly list our assumptions and what they imply.

Our physical assumptions are:

- (i) A classical system has an associated real, differentiable, second-countable finite-dimensional n-manifold Q, called its "configuration manifold".

  The other physical assumptions of classical mechanics offer a prescription for constructing Q in any given case, but we don't care about that here. All we need to know is that Q exists and can be provided to us for any system under consideration.

- (ii) The full state space of a classical system with configuration manifold Q is $\Omega \equiv T^*Q$.

- (iii) A classical system has a suitably differentiable Hamiltonian $H : T^*Q \to \mathbb{R}$ that tells us the energy of each state.

  As with Q, the other physical assumptions provide a prescription for constructing H in any given case, but that doesn't concern us here. We only care that H exists and is known.

These imply the following:

- Q and T^*Q are differentiable manifolds, so we can assume smoothness without any loss of generality. [See the embedding theorem 2.4.]
- Given any open interval $\delta_E \equiv (E - \epsilon, E + \epsilon)$, $\Omega_E \equiv H^{-1}(\delta_E)$ is open in Ω since H is continuous.
- Being an open subset of Ω, Ω_E is an embedded $2n$-submanifold of Ω. [See proposition 2.6.]
- Being manifolds, Q, Ω, and Ω_E are locally compact, paracompact, and Hausdorff. However, being a vector bundle, Ω is always noncompact. Ω_E may or may not be compact. [See section 3.2.]
- Since Q is second-countable, its cotangent bundle T^*Q is second-countable, and the submanifold $\Omega_E \subset T^*Q$ is second-countable. [See the comment in section 6.2 and proposition 2.2.]
- Being a cotangent bundle, Ω has a tautological 2-form ω. [See section 7.6.]
- (Ω, ω) is a symplectic manifold. [See section 7.6.]
- As a manifold that admits a symplectic structure, Ω is orientable and thus admits volume forms. [See section 7.3.]
- ω induces a natural class of volume forms on Ω, given by $\nu = c \cdot \omega^n$ for any $c \neq 0$. [See section 7.3.]

  As discussed earlier, for either sign of c, there exists a unique orientation that makes ν positive on all components. All differential-geometric integrals involving ν implicitly are implicitly in this orientation. The scale of c washes away in the normalization of the resulting measure.

- Any volume form on Ω restricts to a volume form on the $2n$-submanifold Ω_E, so our choice of ν on Ω is inherited by each Ω_E. [See proposition 2.7.]
- We can integrate $\int_\Omega f\nu$ for any real continuous function f on Ω, and we can integrate $\int_O f\nu$ for any open $O \subset \Omega_E$ and real continuous function f on O. [See section 2.3.1.]
- ν defines a positive functional $\hat{\mu} : C^0(\Omega) \to \mathbb{R}^*$ via $\hat{\mu}(f) \equiv \int_\Omega f\nu$. It is a positive linear functional on $C_C(\Omega)$, the vector space of continuous real functions with compact support. [See section 6.1.]
- The Riesz-Markov-Kakutani theorem tells us that $\hat{\mu}$ induces a unique almost-Radon measure μ whose measure-theoretic integral $\int_\Omega f\mu$ is consistent with $\hat{\mu}$ (and thus equals the differential-geometric integral $\int_\Omega f\nu$) for all $f \in C_C(\Omega)$. This measure μ is on some σ-algebra containing the Borel algebra of Ω, and on which it is complete. [See section 6.2.]
- The induced μ restricts to an almost-Radon measure on the Borel algebra of Ω. [See proposition 4.9.]
- The Borel algebra of Ω_E is a subset (but not σ-subalgebra) of the Borel algebra of Ω. We'll refer to these as B_{Ω_E} and B_Ω. [See proposition 4.8, part (i).]
- μ restricts to an almost-Radon measure on B_{Ω_E}. This extends to their measure-completions. [See proposition 4.8, parts (ii) and (iii), which apply equally well to almost-Radon measures. For the completions, see proposition 4.14.]

  Note that we will *not* be seeking a finite measure on B_Ω, but rather on B_{Ω_E}. We could perform the above steps on Ω_E directly: restricting the volume form to Ω_E, constructing a positive linear functional from it, applying the Riesz-Markov-Kakutani theorem to that, and restricting the result to B_{Ω_E} to get an almost-Radon measure. The result would be the same. See proposition 6.2. As will soon become apparent, there is a benefit to working with Ω and only restricting at the end.

- For a finite measure, Radon and almost-Radon are the same. Any measure which can serve for us must have this quality. [See the notes under the Riesz theorem in section 6.2.]

- Any finite measure on a locally compact, second-countable, Hausdorff space is regular and locally finite, so our almost-Radon measure μ is the only possible candidate for a finite measure on B_{Ω_E}. [See proposition 6.3.]
- If μ is finite on B_{Ω_E}, then it normalizes into a probability measure.

  This renders irrelevant the choice of scale in $\nu = c \cdot \omega^n$.
- Any finite (and nonzero) measure μ on B_{Ω_E} defines a uniform probability field on Ω_E — namely the normalized $P = \mu/\mu(\Omega_E)$. [See section 5.1.]

  P trivially extends to a probability measure on the completion, since the completion traffics in measure-zero changes.

  Bear in mind that $\mu(\Omega_E) > 0$ always. Open sets are big. They are embedded n-submanifolds, and a volume form always produces a nonzero volume on one.

The only unresolved question is whether μ on B_{Ω_E} is finite. Without some sort of additional assumption, this cannot be guaranteed. Worse, if our induced μ is not finite then *no* suitable probability measure exists. It isn't simply a matter of digging harder for one in a system-specific way.

We've already seen that T^*Q can't be compact but Ω_E may or may not be. We don't need Ω_E to actually be compact in order for μ to be finite. A sufficient (but not necessary) condition is that Ω_E be contained in a compact subset of Ω.

Our measure μ on B_Ω is locally finite. Since Ω is locally compact and Hausdorff, every compact set is of finite measure. By monotonocity, $\mu(\Omega_E) \leq \mu(K)$ if $\Omega_E \subseteq K$. If K is compact (and thus has finite measure), Ω_E must too.

Note that we can meaningfully define this criterion because μ on B_{Ω_E} is the restriction of μ on B_Ω. This is why we chose to take that route. If we had constructed μ directly on B_{Ω_E} from the outset, we couldn't speak of the measure of sets larger than Ω_E. We have a great deal of information about Ω and its relationship to Ω_E that we would be discarding, and which is of great utility for our current purpose.

The form of H offers some justification for such an assumption. Hamiltonians usually are quadratic in the momenta, but we don't even need to be that specific. All we must do is ensure that $H^{-1}(\delta_E)$ is confined to a finite volume of the cotangent space. This will be the case if H scales to $+\infty$ with the magnitude of each momentum coordinate. To keep to δ_E, we then must stay within a bounded range of momenta.

However, such a constraint on the momenta alone does not guarantee that Ω_E is compact. If Q itself is compact then, under such a scaling constraint, Ω_E sits inside a compact subset of T^*Q and thus has finite volume. However, in general Q need not be compact.

If Q is noncompact but H scales to $+\infty$ in each coordinate direction, then — for the same reason as with the momenta — Ω_E confines us to a bounded domain, and Ω_E will sit in a compact subset of Ω and thus have finite volume. This is the case when the spatial volume is explicitly bounded (ex. a gas in a box) or there is an external potential which grows to $+\infty$ in all directions. As long as any unbounded coordinates cause the energy to scale to $+\infty$, we are fine. However, if we have an unbounded degree of freedom that is a symmetry of H, or if we have a negative potential as we grow the q's, then we're out of luck.

This isn't farfetched. Ex. a gas in an unbounded volume with no external potential would lead Ω_E to have infinite volume.

We therefore must add another physical assumption, something to the effect that either Q is compact or the Hamiltonian scales to infinity with every unbounded variable. This isn't very rigorous as stated, so let's frame it more formally:

- (iv) For every E, there exists some $\epsilon > 0$ s.t. $H^{-1}((E-\epsilon, E+\epsilon))$ is contained in a compact subset of Ω.

  Note that if this holds for ϵ, it holds for every $\epsilon' < \epsilon$ too, so we still are free to choose each δ_E to be as small as needed.

Under assumptions (i)-(iv), we thus have a natural notion of "uniform probabilities" on each Ω_E.

More precisely, we take as given assumptions (i)-(iii) and then either contrive our system to satisfy (iv) or must verify that it does.

If our system does not satisfy (iv), then it is still possible that the almost-Radon μ produced by the Riesz theorem is finite. In that case, we must manually confirm this.

References

[1] Brian Conrad. *Differential Geometry Handouts*. `https://virtualmath1.stanford.edu/~conrad/diffgeomPage/handouts/paracompact.pdf`.

[2] Lynn Arthur Steen and Jr. J. Arthur Seebach. *Counterexamples in Topology*. Springer Verlag, 2nd edition, 1978.

[3] J.R. Munkres. *Topology*. Prentice Hall, 2nd edition, 2000.

[4] R. Abraham, J.E. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications*. Springer-Verlag, 2nd edition, 1988.

[5] Morris W. Hirsch. *Differential Topology*. Springer-Verlag, 1976.

[6] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[7] John M. Lee. *Introduction to Smooth Manifolds*. Springer, 2nd edition, 2013.

[8] `https://math.stackexchange.com/questions/527642/the-equivalence-between-paracompactness-and-second-countablity-in-a-locally-eucl`.

[9] James R. Munkres. *Analysis on Manifolds*. Addison-Wesley, 1991.

[10] M. Reed and B. Simon. *Methods of Modern Mathematical Physics, IV: Analysis of Operators*. Academic Press, 1978.

[11] Terence Tao. *Analysis II*. Hindustan Book Agency, 3rd edition, 2014.

[12] Allan Gut. *Probability: A Graduate Course*. Springer, 2005.

[13] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, 1970.

[14] `https://mathoverflow.net/questions/343636/is-every-finite-borel-measure-on-a-locally-compact-hausdorff-sigma-compact-a`.