

Chapter 1

Ensemble spaces

Here we summarize the work in progress for a general theory of states and processes that is applicable to any physical system. The tentative core concept is the ensemble, therefore we want to find definitional requirements for ensembles and then find suitable assumptions to recover the different theories (e.g. classical, quantum and thermodynamics).

Conceptually, we want to argue that physical theories are, and can only be, about ensembles. At a practical level, most of the time we can only prepare and measure statistical properties as we do not have perfect control over any system. The cases where properties can be prepared with one hundred percent reliability can still be understood as ensembles of identical preparations, though one can argue for an ensemble of external conditions or internal dynamics. At a conceptual level, the goal of physics is to write laws that apply all the time: every time that one prepares a system according to a particular procedure, lets it evolve in particular conditions, he will obtain a particular result. That is, the idea of repeatability of experimental results implicitly assumes that the objects of scientific inquiry are not single instances, but the infinite collections of all reproducible instances.

1.1 Ensembles

TODO: Preamble that argues the justification with physics examples.

Axiom 1.1 (Axiom of ensembles). *The state of a system is represented by an **ensemble**, which represents all possible preparations of equivalent systems prepared according to the same procedure. The set of all possible ensembles for a particular system is an **ensemble space**. Formally, an ensemble space is a T_0 second countable topological space where each element is called an ensemble.*

Justification. In physics, ensembles are usually defined as probability distributions over states. This is conceptually problematic. Any actual physical preparation is always associated with some uncertainty. That is, the system will never be replicated exactly. This is true even in quantum mechanics: while we can always prepare a hydrogen atom in the ground state, we can never prepare an atom in the same exact position with the same exact velocity. The ability to prepare ensembles in which the position and the excitation of the atom can be assumed to be uncorrelated is what allows us to concentrate on one of the properties. Therefore, ensembles are the only thing that can actually be prepared in prac-

tice. Pure states, then, should not be taken as primitive notions but should be understood as an idealized ensemble.

From a more conceptual level, reproducibility is baked into the requirement of experimental verifiability. A physical law, then, must be understood as describing a relationship that always exist whenever the same set of circumstances is replicated. Given that we need to always be able to replicate those circumstances “one more time”, the relationship is about countably infinite similar preparations: an ensemble. Therefore, to the extent that physics is about reproducible experimental results, the basic description of a system is in terms of ensembles. This justifies the use of ensembles as the fundamental object to describe the state of a system.^a

Ensembles are experimentally defined objects, and therefore they are possibilities of an experimental domain. Therefore an ensemble space is a T_0 second countable topological space where each element is an ensemble and the topology is induced by the verifiable statements. $\square$

^aNote that reproducibility also already implies that all properties that characterize an ensemble must be relative to the procedure. If the properties depended, for example, on absolute space or absolute time, then different practitioners would not be able to prepare the same ensemble.

A general theory of ensembles will need to be able to recover classical and quantum spaces under appropriate physically meaningful assumptions. Here we review these cases and give the appropriate mathematical definitions, highlighting potential issues, as these serve as models and examples for a general theory.

There are two aspects that are common to all cases and will be the target of generalization: statistical mixing (captured by a convex structure) and entropy (captured by a strictly concave function). Note: all logs are assumed in base 2.

Discrete classical theory

Definition 1.2. *The ensemble space of a **discrete classical theory** is a simplex with the standard convex structure and the Shannon entropy. The simplex can have finitely or countably many vertices.*

That is, a discrete classical ensemble space is a set $\mathcal{E}$ containing n points $\{s_i\}_{i=1}^n$ and for which every element $\mathbf{e} \in \mathcal{E}$ is characterized by a unique convex combination $\mathbf{e} = \sum_i p_i s_i$ such that $\sum_i p_i = 1$. The space is closed under convex combinations of finitely many elements. The entropy is given by $S(\mathbf{e}) = -\sum_i p_i \log p_i$.

Remark. The correct closure under convex combinations of countably many elements is not clear.

Finite case

The space of classical distributions over a discrete space corresponds to a simplex (<https://en.wikipedia.org/wiki/Simplex>). In the finite case, the pure states $\mathcal{S} = \{s_i\}_{i=1}^n \subset \mathcal{E}$ are finitely many and each ensemble $\mathbf{e} = \sum p_i s_i$ is uniquely identified by a decomposition of pure states. Effectively, each ensemble is a probability distribution over the pure states. Mathematically, each point of the space is a convex combination of the vertices. The simplex has a center point, which corresponds to the maximally mixed state, a uniform distribution over all pure states.

The entropy is given by the Shannon entropy $-\sum_i p_i \log p_i$. This means that the entropy of each pure state is zero and the entropy of the maximally mixed state is $\log n$ where n is the number of pure states. The entropy increases as we go from pure states to the maximally mixed state. The level sets (i.e. the fibers) of the entropy form a series of concentric “shells”.

The fact that all pure states have the same entropy is an additional hypothesis that cannot work in general. To see this, consider the case where the state is defined by the number of molecules for two substances. This space is the product of two independent variables n_a and n_b . If we have a uniform distribution over N_a cases of n_a and N_b cases of n_b , the total number of cases is $N_a N_b$. Therefore the entropy of the joint state is the sum of the entropy of the marginals. However, if we pair n_a with the total number of molecules $n_{(a+b)}$ we have a problem. The issue is that variable $n_{(a+b)}$ corresponds to a variable number of joint cases. Therefore the case where the ensemble space is a simplex but the entropy is not the Shannon entropy (e.g. is the Shannon entropy plus the contributions of entropy from each vertex) is a physically meaningful case that should be possible in the general theory.

Countable case

The countable case is not well-defined mathematically.

The obvious extension is to include all series $\{p_i\} \in [0, 1]$ that converge to one. That is, the space of all probability measures over a countable discrete space. Since we cannot create a uniform distribution over infinitely many cases, there is no center point. Effectively, there is a “hole” in the middle. More precisely, the space is not topologically closed in the sense it does not contain all the limit points.

[TODO] It may be useful to characterize this “hole” and the limit points. There should be at least one limit point for each series $\{p_i\}$ that converges to a finite $p < 1$. Intuitively, we can keep that part of the distribution constant while we spread the rest uniformly to all cases. Each should reach a different limit point.

However, the space of all probability measures is too large. Note that the entropy is not finite for all convergent p_i (i.e. infinite convex combinations). For details, see <https://arxiv.org/pdf/1212.5630.pdf>. Given that we want the entropy to exist and be finite for all ensembles, this generalization (like in <https://ncatlab.org/nlab/show/superconvex+space>) does not seem physically warranted.

Also note that expectation values are not guaranteed to be finite either, and requiring a particular observable to be finite further restricts the space. This restriction may be desirable for another reason: a discrete ensemble space has no notion of the ordering of the pure states. Physically, this would mean that the states with 1, 100, or 1 trillion particles are “equally distant” (i.e. all infinite permutations are allowed). Requiring the expectation of the number of particles to be finite (e.g. $\sum_i N(i)p_i < \infty$) should effectively encode the infinite ordering in the rate of convergence of the probability distributions (i.e. not all infinite permutations would be allowed).

No uncountable case

The uncountably infinite case is not physically relevant (i.e. no second countable discrete topology, cases are not experimentally decidable). Also note that any set of real numbers whose sum is finite can have only countably many non-zero elements. To understand why, note that there can only be finitely many terms above any particular positive value if their sum is to remain finite. Effectively, the uncountable case would be stitching together infinitely

many countable cases.

Continuous classical theory

Definition 1.3. *The ensemble space of a **continuous classical theory** is the convex space given by the probability distributions (i.e. probability measures) over phase space (i.e. a symplectic manifold) that allow a continuous probability density (i.e. the Radon-Nikodym derivative exists - the probability measures are absolutely continuous with respect to the Liouville measure - and is continuous). Formally, let (X, ω) be a symplectic manifold. The ensemble space is given by $\mathcal{E} = \{\rho \in C(X) \mid \int_X \rho d\mu = 1\}$. The entropy is given by the Shannon/Gibbs entropy calculated using the probability density (i.e. the Radon-Nikodym derivative between the probability measure and the Liouville measure). That is, $S(\rho) = - \int_X \rho \log \rho d\mu$.*

In the continuous case, the space of ensembles is the space of continuous integrable functions over a symplectic manifold (e.g. over phase space) that integrate to one. That is, if X is a symplectic manifold, then $\mathcal{E} = \{\rho \in C(X) \mid \int_X \rho(x) d\mu = 1\}$ where $\mu = \int \omega^n$ is the Liouville measure. The space does not include the pure states, as they correspond to delta distributions $\mathcal{S} = \{\delta_x\}_{x \in X}$: they are limit points. In some sense, a distribution can be understood as a convex integral of pure states: $\rho(x) = \int_X \rho(y) \delta_x(y) d\mu$. Even in this case, each ensemble can be understood as a unique probability distribution (density in this case) over all pure states. The symplectic nature of the manifold is required to assign a frame invariant density to states and a frame invariant notion of independence between DOFs, as we saw in the classical mechanics section of reverse physics.

The fact that pure states are not part of the space, but limits, means we need a more sophisticated definition of pure states than simply the extreme points of the convex space of ensembles.

The entropy is given by $-\int \rho \log \rho d\mu$ where μ is the Liouville measure and ρ is the probability density over canonical coordinates. If a different measure is used, or if the coordinates are not canonical, the formula gives the wrong result.

Similarly to the infinite discrete case, the entropy can be infinite and expectation values can be infinite. The added complication is the frame invariance: it would not make sense to have the expectation for position in one frame to be finite while infinite in another. Requiring all functions of position/momentum to have finite expectation restricts the distributions to those with finite support. Requiring all polynomial functions of position/momentum to have finite expectation restricts the distributions to those that decay faster than any polynomial.

Unlike the discrete classical case, subspaces and dimensionality of subspaces cannot be defined without the entropy. The issue is that we need a measure on the set of pure states, and the convex structure cannot provide it. The entropy, however, does as the supremum of the entropy for all distributions with support U is $\log \mu(U)$. As we will see, the entropy can be used to both identify subspaces and recover the Liouville measure.

Note that we exclude discontinuous distributions as starting from the space of discontinuous distributions does not allow one to recover the topology of the base space. Suppose, in fact, to have the space of all distributions over the real line that integrate to one. Now imagine switching the interval $[0, 1)$ with the interval $[1, 2)$. This is an isomorphism on the space of discontinuous distributions. Therefore given the space of distributions, we cannot recover the topology of the real line. Now take the space of continuous functions that integrate to one. Take those with support $[0, 1) \cup [2, 3)$. This is the set of convex combinations of the distri-

butions with support $[0, 1)$ and $[2, 3)$ precisely because the sets are not contiguous and the distributions will have to go to zero at the boundary. However, the distributions with support $[0, 2)$ are not just the convex combinations of those with support $[0, 1)$ and $[1, 2)$ because this will include the continuous functions that have a non-zero value at 1.

[TODO] It may be interesting to study the shell of zero entropy states. For example, it should not be path connected: all uniform distributions with support of the same finite size (in terms of the Liouville measure) will have the same entropy. The region, however, need not be contiguous. Since we cannot continuously transform a single region into two disjoint regions, there will be different distributions at zero entropy that cannot be transformed continuously.

Quantum theory

Definition 1.4. *The ensemble space of a **quantum theory** is modeled by the convex space given by the density operators (i.e. positive semi-definite self-adjoint operators with trace one) of a Hilbert space equipped with the von Neumann entropy.*

That is, given a Hilbert space $\mathcal{H}$, the ensemble space is the space of positive semi-definite self-adjoint operators with trace one $M(\mathcal{H})$. The space of pure states is given by the projective space $P(\mathcal{H})$. The entropy of an ensemble $\rho \in M(\mathcal{H})$ is given by the von Neumann entropy $S(\rho) = -\text{tr}(\rho \log \rho)$.

Finite dimensional case

The simplest non-trivial case is the qubit, for which the Bloch ball is the space of ensembles $M(\mathcal{H})$. The interior of the Bloch ball corresponds to mixtures while the surface corresponds to the pure states $\mathcal{S} = P(\mathcal{H}) = \{|\psi\rangle\langle\psi| \}_{\psi \in \mathcal{H}}$. In quantum ensemble spaces there is no unique decomposition. Note that the space is exactly characterized by knowing which different mixtures provide the same ensemble.

The multiple decompositions make the ensemble space behave in a way that is a hybrid between the classical discrete and continuous. Pure states are properly a part of the ensemble space, as in the discrete case, and we can describe each mixture in terms of finitely many pure states. However, the pure states are a continuum, therefore we can also define probability densities over the space, convex integrals. For example, for a single qubit, the maximally mixed state (the center of the ball) can be equally described as the equal mixture of two opposite states (e.g. spin up and spin down, or spin left and spin right). However, it can also be described as the equal mixture of the whole sphere.

Note that complex projective spaces are symplectic, which is what allows one to define frame invariant densities. The goal is to have one argument applied to the generic definition as to why the space of pure states must be symplectic. Also note that the two dimensional sphere is the only symplectic sphere. By homogeneity, we should be able to argue that the space is symmetric around the maximally mixed state, and is therefore a sphere. The symplectic requirement would select dimension two. Note that real and quaternionic spaces would be excluded by this argument.

The von Neumann entropy for the maximally mixed state is $\log n$ where n is the dimensionality of the Hilbert space. Again we see that the maximum entropy gives us a measure of the size of the space. Note that, to calculate the von Neumann entropy, we are diagonalizing the density matrix ρ . This means finding a set of orthogonal pure states s_i such that $\rho = \sum p_i s_i$ is a convex combination. Note that the convex hull of a set of n orthogonal pure states is an n -dimensional simplex whose center is the maximally mixed state. Therefore, we are looking

for a simplex that contains ρ and the maximally mixed state. In the two dimensional case, ρ is an interior point of the Bloch ball. Take the line that connects ρ to the center. The two points of the sphere are the extreme points for the decomposition. The distance from the points will be proportional to the probability. Because of this property, the von Neumann entropy is the smallest Shannon entropy between all possible decompositions.

Countably infinite dimensional case

The countable infinite dimensional case presents similar problems as the classical case, and adds others. As in the classical infinite cases, the maximally mixed state (i.e. uniform distribution) is not in the convex space and the entropy is not finite for all infinite convex combinations. As in the classical continuous case, there is the issue of finite expectation of position/momentum in all frames. The problem is compounded by the fact that one cannot require finite expectation for all functions of position and momentum: finite support in position automatically implies infinite support on momentum, since the distribution in momentum is the Fourier transform of that in position.

The Hilbert space for a discrete variable with infinite range (e.g. number of particles) and a continuous variable (e.g. position/momentum) is the same. The first is defined as the space of square convergent complex sequences l^2 while the second is the space of square integrable complex functions L^2 . Given that L^2 allows a countable basis, the two are isomorphic. This also means that all spaces with finitely many degrees of freedom are also isomorphic. This makes the problem of infinite expectations even more problematic.

Note that Schwartz spaces have finite expectation for all polynomial functions of position and momentum. Given that infinite permutations can change the rate of convergence, the Schwartz space has an idea of what is further away from the origin, unlike Hilbert spaces. [TODO] It is not yet clear to us whether the Schwartz space for one DOF is the same as the one for two DOFs.

1.2 Mixtures and convex spaces

Definition 1.5. Given a real number $p \in [0, 1]$, its complement is defined as $\bar{p} = 1 - p$.

Axiom 1.6 (Axiom of mixture). An ensemble space $\mathcal{E}$ is equipped with an operation $+$: $[0, 1] \times \mathcal{E} \times \mathcal{E} \rightarrow \mathcal{E}$ called **mixing**, noted with the infix notation $pe_1 + \bar{p}e_2$, with the following properties:

- **Continuity:** $pe_1 + \bar{p}e_2$ is continuous in all its arguments (i.e. p , e_1 and e_2)
- **Identity:** $1e_1 + 0e_2 = e_1$
- **Idempotence:** $pe_1 + \bar{p}e_1 = e_1$ for all $p \in [0, 1]$
- **Commutativity:** $pe_1 + \bar{p}e_2 = \bar{p}e_2 + pe_1$ for all $p \in [0, 1]$
- **Associativity:** $p_1e_1 + \bar{p}_1 \left(\left(\frac{p_3}{p_1} \right) e_2 + \frac{p_3}{p_1} e_3 \right) = \bar{p}_3 \left(\frac{p_1}{p_3} e_1 + \left(\frac{p_1}{p_3} \right) e_2 \right) + p_3 e_3$ for all $p_1, p_3 \in [0, 1]$

Justification. This axiom captures the ability to create a mixture merely by selecting between the output of different processes.

Given that mixing represents an experimental relationship, and all experimental relationships must be continuous in the natural topology, mixing must be a continuous function. Note that p is a continuous quantity, and therefore the natural topology is the one of the reals.

Let $\mathbf{e}_1$ and $\mathbf{e}_2$ be the ensembles that represent the output of two different processes P_1 and P_2 . Let a selector S_p be a process that outputs two symbols, the first with probability p and the second with probability $\bar{p}$. Then we can create another process P that, depending on the selector, outputs either the output of P_1 or P_2 . Therefore we are justified in assuming the mixing operation among ensembles.

If $p = 1$, the output of P will always be the output of P_1 . This justifies the identity. If P_1 and P_2 are the same process, then, again, the output of P will always be the output of P_1 . This justifies the idempotence. As long as the same probability is matched to the same output, the process P is identical. This justifies commutativity. If we are mixing three processes P_1 , P_2 and P_3 , as long as the final probabilities are the same, it does not matter if we mix P_1 and P_2 first or P_2 and P_3 . This justifies associativity. $\square$

Remark. Given symmetry and associativity, we can write $p_1\mathbf{e}_1 + p_2\mathbf{e}_2 + p_3\mathbf{e}_3$ where $p_2 = \bar{p}_1\left(\frac{p_3}{p_1}\right) = \bar{p}_3\left(\frac{p_1}{p_3}\right) = 1 - p_1 - p_3 = \overline{(p_1 + p_3)}$. We can also extend to mixtures of finitely many elements $\sum p_i\mathbf{e}_i$ where $\sum_i p_i = 1$.

Corollary 1.7. *An ensemble space is a convex space.*

Remark. Basic definitions for convex spaces are taken from <https://ncatlab.org/nlab/show/convex+space> and <https://arxiv.org/abs/0903.5522>. The notation and terminology will be slightly different to better map to physics ideas.

Definition 1.8. *Let $\mathcal{E}$ be an ensemble space. Let $\rho = \sum_i p_i\mathbf{e}_i$ where $\rho, \{\mathbf{e}_i\} \in \mathcal{E}$ and $p_i \in (0, 1]$ such that $\sum p_i = 1$. We say that ρ is a **mixture** of $\{\mathbf{e}_i\}$ and each $\mathbf{e}_i$ is a **component** of ρ .*

Remark. In terms of the convex space, all the mixtures between two elements corresponds to the segment between them; all the mixtures between three elements correspond to the triangle formed by the three elements and so on. An element $\mathbf{e}_1$ is a component of a different element $\mathbf{e}_2$ if the segment can be continued on the side of $\mathbf{e}_2$. Note that two elements can be a component of each other. If two elements are not a component of each other, then they are the extreme points of the line that connect the two.

Definition 1.9. *Let $\mathcal{E}$ be an ensemble space and $\mathbf{e}_1, \mathbf{e}_2 \in \mathcal{E}$. We say that they **have a common component** if we can find $\mathbf{e}_3 \in \mathcal{E}$ such that $\mathbf{e}_1 = p_1\mathbf{e}_3 + \bar{p}_1\mathbf{e}_4$ and $\mathbf{e}_2 = p_2\mathbf{e}_3 + \bar{p}_2\mathbf{e}_5$ for some $\mathbf{e}_4, \mathbf{e}_5 \in \mathcal{E}$ and $p_1, p_2 \in (0, 1)$. They are **distinct**, noted $\mathbf{e}_1 \perp \mathbf{e}_2$, otherwise.*

Remark. In terms of convex spaces, two distinct ensembles are always the extreme points of the line that connects them. Additionally, for any other ensemble $\mathbf{e}$, one of the two ensembles must be the extreme point of the line that connects it to $\mathbf{e}$.

Corollary 1.10. *The previous definitions obey the following:*

1. every ensemble has a common component with itself, therefore every ensemble is not distinct from itself
2. distinctness is a irreflexive symmetric relation
3. if $\mathbf{e}_1$ is a component of $\mathbf{e}_2$, then $\mathbf{e}_1$ and $\mathbf{e}_2$ have a common component

Proof. 1. Since by idempotence $\mathbf{e} = p\mathbf{e} + \bar{p}\mathbf{e}$ for any p , we can satisfy the definition by setting $\mathbf{e}_1 = \mathbf{e}_2 = \mathbf{e}_3 = \mathbf{e}_4 = \mathbf{e}_5 = \mathbf{e}$. Therefore every ensemble has a common component with itself and every ensemble is not distinct from itself.

2. The previous property shows that distinctness is irreflexive. The definition of common component is symmetric and therefore so is distinctness.

3. By idempotence, we can write $e_1 = p_1 e_1 + \bar{p}_1 e_1$ for some $p_1 \in (0, 1)$. Since e_1 is a component of e_2 , we can write $e_2 = p_2 e_1 + \bar{p}_2 e_3$ for some $p_2 \in (0, 1)$ and $e_3 \in \mathcal{E}$. Therefore e_1 and e_2 have a common component. $\square$

Conjecture 1.11. *Let $e, e_1, e_2 \in \mathcal{E}$. If e is distinct from a mixture of e_1 and e_2 then it is distinct from all mixtures of e_1 and e_2 .*

Remark. Note that if $e \perp p e_1 + \bar{p} e_2$ for some p , it is not necessarily true that $e \perp e_1$ and $e \perp e_2$. Conversely, if even if $e \perp e_1$ and $e \perp e_2$, it is not necessarily distinct from all mixtures of e_1 and e_2 .

For a physical example, consider the Bloch sphere and the states x^+, x^-, z^+ , and z^- . These are all distinct ensembles. We have $\frac{1}{2}x^+ + \frac{1}{2}x^- = \frac{1}{2}z^+ + \frac{1}{2}z^-$. Therefore a mixture of x^+ and x^- has a common component with z^+ .

Definition 1.12. *Let $\rho = p e_1 + \bar{p} e_2$ with $p \in (0, 1)$. Then we say that e_2 is a **p -complement** of e_1 towards ρ . An ensemble space is **complemented** if all p -complements are unique for all $p \in (0, 1)$.*

Proposition 1.13. *An ensemble space is complemented if and only if it is a convex subset of a real vector space for which mixtures are linear combination.*

Proof. Theorem 4 in <https://arxiv.org/abs/1105.1270> states that a convex space embeds into a real vector space with $c_\lambda(x, y) = \lambda x + \bar{\lambda} y$ if and only if

$$c_\lambda(x, y) = c_\lambda(x, z) \quad \forall \lambda \in (0, 1) \implies y = z.$$

This is exactly the condition that the λ -complement is unique. $\square$

Proposition 1.14. *Discrete classical ensemble spaces, continuous classical ensemble spaces and quantum ensemble spaces satisfy the axiom of mixture and are complemented ensemble spaces.*

Proof. The space of probability measures, discrete or continuous, $\mathcal{E}$ is a convex subset of the vector space of finite signed measures. This means that it is closed under convex combinations: $\sum p_i e_i$ for all $\{e_i\} \subseteq \mathcal{E}$ and $\sum p_i = 1$. The properties of mixing are inherited from the properties of linear combinations. Therefore the discrete and continuous classical ensemble spaces satisfy the axiom of mixture. Moreover, $\mathcal{E}$ is a convex subset of a vector space, it is complemented by proposition 1.13.

Similarly, the space of positive semi-definite self-adjoint operators with trace one is a convex subset of the vector space of self-adjoint operators. Therefore it is closed under convex combinations and it will satisfy the axiom of mixture. Moreover, it is complemented convex by proposition 1.13. $\square$

Remark. Since all current physical ensemble spaces are complemented, it is not yet clear whether an ensemble space must be complemented or not. It is difficult to justify as it is a claim on existence of more refined ensembles. It may very well be that this is just an assumption on the decomposition, which may fail in new physical conditions (e.g. at Planck scale). It may also be something that the strict concavity of the entropy requires.

1.3 Entropy

Definition 1.15. Given the coefficients $\{p_i\} \in [0, 1]$ such that $\sum p_i = 1$, the **entropy of the coefficients** (also known as Shannon entropy) is defined as $I(\{p_i\}) = -\sum p_i \log p_i$.

Axiom 1.16 (Axiom of entropy). An ensemble space $\mathcal{E}$ is equipped with a function $S : \mathcal{E} \rightarrow \mathbb{R}$ called **entropy** with the following properties

- **Continuity**
- **Strict concavity:** $S(p_1 e_1 + p_2 e_2) \geq p_1 S(e_1) + p_2 S(e_2)$ with the equality holding if and only if $e_1 = e_2$
- **Upper variability bound:** $S(p_1 e_1 + p_2 e_2) \leq I(p_1, p_2) + p_1 S(e_1) + p_2 S(e_2)$; if the equality hold then, e_1 and e_2 are **disjunct**, noted $e_1 \perp\!\!\!\perp e_2$

Justification. The entropy quantifies the variability of the elements within the ensemble. This justifies the existence of the entropy.

Small changes in the ensemble should produce small changes in the variability, which justifies continuity. Mixing different ensembles will increase the variability, which justifies concavity. Mixing the same ensemble, however, should not increase the variability, which justifies strict concavity.

The variability increase is maximal when the two ensembles are disjunct, meaning they have no elements in common. In this case, the increase in variability is given only by the coefficients and must follow the Shannon entropy as it is the only continuous indicator of variability that is linear in probability. This justifies the upper bound. $\square$

Remark. In <https://arxiv.org/abs/1912.02012> we showed how the Shannon entropy quantifies the variability of elements within a distribution. Is it possible to rederive I from requirements on convex spaces instead of distributions in particular? This would make the axiom less ad-hoc.

Axiom 1.17 (Axiom of entropic disjunctness). The entropy function of an ensemble space obeys the following properties

- **Disjunctness implies distinctness:** $e_1 \perp\!\!\!\perp e_2$ implies $e_1 \perp e_2$
- **Mixtures preserve disjunctness:** let $e_1 = p e_2 + \bar{p} e_3$, then $e_4 \perp\!\!\!\perp e_1$ if and only if $e_4 \perp\!\!\!\perp e_2$ and $e_4 \perp\!\!\!\perp e_3$

Justification. If two ensembles are disjunct, they have no elements in common. Therefore there will be no common component with the two. We are justified to assume that disjunctness implies distinctness.

Suppose an ensemble e_1 is disjunct from the mixture of two ensembles e_2 and e_3 . That it has no elements in common with the mixture, which means it has no elements in common with either of them. Therefore e_1 is disjunct from both e_2 and e_3 . Now suppose e_1 is disjunct from both e_2 and e_3 . That it means it has no elements in common with either of them, and therefore it does not have any elements in common with their mixture. Therefore e_1 is disjunct from the mixture of two ensembles e_2 and e_3 . We are justified to assume that mixtures preserve disjunctness. $\square$

Remark. It is unclear whether or not these properties can be proven by the axiom of mixture. It is possible to prove that two disjunct ensembles are not one the component of the other. The entropy of the coefficients (i.e. the Shannon entropy) has a vertical asymptote at

each end. If two elements are disjunct, the entropy along the segment that connects the two follows that function plus a linear function, which does not change the vertical asymptotes. Therefore the segment cannot be extended while preserving the convexity of the entropy, and therefore the two distinct elements must be the extremes of the line that connects the two. This is necessary for distinctness, but not sufficient. Maybe, extending the argument, one can at least prove that disjunctness implies distinctness.

Proposition 1.18. *The discrete classical ensemble spaces, continuous classical ensemble spaces and quantum ensemble spaces satisfies the axiom of mixture and the axiom of entropic disjunctness.*

Proof. In both classical cases, the entropy is a function of only the probabilities of the decomposition in terms of pure states. The function is continuous in terms of those variables as $x \log x$ is a continuous function. [TODO find short proof of convexity and maximum bound]

In the quantum case, the entropy is a continuous function of the density operator as the trace, multiplication and logarithm of an operator are all continuous functions. [TODO find short proof of convexity and maximum bound] $\square$

Remark. There should be an interplay between convexity and entropy which limits either the space or the possible entropy. For example, consider a line. It is a convex space with no extreme points. We can parameterize the space with a variable x such that the mixture of two ensembles is identified by the average x . Entropy can be written as $S(x)$. Since $S(x)$ is strictly concave over an infinite range, it will tend to minus infinity either when x tends to infinity or minus infinity. Now take a line that intersects the curve. This will intersect it at only two points. By strict convexity, there is only going to be a single point with maximum vertical distance to the line, which has to be greater than zero. Now imagine moving the line down by one vertical unit. The vertical distance now is greater than one. This means that the average entropy can increase more than one during mixture, which violates the upper variability bound. Therefore a line cannot be an ensemble space.

Note that an open segment is absolutely fine. We can, in fact, imagine a two state discrete classical ensemble space with the endpoint removed. This suggests that the convex space knows something about the geometry (probably parallelism and ratio of length - i.e. the affine structure).

Conjecture 1.19. *Let $\mathbf{e}_1, \mathbf{e}_2 \in \mathcal{E}$ be such that $p\mathbf{e}_1 + \bar{p}\mathbf{e}_2 = p\mathbf{e} + \bar{p}\mathbf{e}_2$ for some $p \in (0, 1)$ and $\mathbf{e} \in \mathcal{E}$. Then $\mathbf{e}_1$ and $\mathbf{e}_2$ are not disjunct.*

Remark. It seems very unlikely that differences in states that can be obscured by mixing would correspond to disjunct states. The more general question is whether the strict concavity of the entropy allows non-complemented spaces.

Definition 1.20. *Two sets of ensembles $U, V \subseteq \mathcal{E}$ are disjunct if $\mathbf{e}_1 \perp \mathbf{e}_2$ for all $\mathbf{e}_1 \in U$ and $\mathbf{e}_2 \in V$.*

Definition 1.21. *An ensemble space $\mathcal{E}$ is **reducible** (or **classical**) if distinct ensembles are also disjunct. That is, $\mathbf{e}_1 \perp \mathbf{e}_2$ implies $\mathbf{e}_1 \perp \mathbf{e}_2$.*

Justification. Given that $\mathbf{e}_1 \perp \mathbf{e}_2$ implies $\mathbf{e}_1 \perp \mathbf{e}_2$, classical spaces add the opposite implication. Therefore they exclude the case where two ensembles are disjunct but not distinct.

This case describes two ensembles that have elements in common, but there is no ensemble corresponding to those common elements. That is, we cannot refine the ensembles into three distinct ones: one with only elements of the first, one with only elements of the second and one with elements of both. In other words, we cannot reduce the coarser description of the system into finer distinct descriptions. In the excluded case, then, the coarser description is irreducible into finer ensembles. This justifies the definition.

As we prove below, that case does not exist in classical mechanics, but exists in quantum mechanics. This justifies calling the property classical. $\square$

Remark. The above property should ultimately be responsible for the single decomposition in terms of pure states.

Proposition 1.22. *Continuous and discrete classical ensemble spaces are reducible.*

Proof. In both cases, ensembles are probability measures: the first over the Borel algebra of a symplectic manifold; the second over the power set of countably many elements. In both cases, the upper bound of the entropy is maximized if and only if the probability measures being combined have disjoint support. This means that two ensembles are disjunct if and only if the respective probability measures have disjoint support. Intuitively, two probability distributions can have a common sub-distribution if and only if they overlap. Therefore two ensembles representing probability measures with disjoint support are exactly ensembles that are distinct. Therefore, in both cases, the ensemble space is reducible. $\square$

1.4 Measures on the ensemble space

Proposition 1.23 (Exponential entropy subadditivity). *Let $\mathbf{e}_1, \mathbf{e}_2 \in \mathcal{E}$. Let $S_1 = S(\mathbf{e}_1)$ and $S_2 = S(\mathbf{e}_2)$. Let $\mathbf{e} = p\mathbf{e}_1 + \bar{p}\mathbf{e}_2$ for some $p \in [0, 1]$ and $S = S(\mathbf{e})$. Then $2^S \leq 2^{S_1} + 2^{S_2}$, with the equality if and only if $\mathbf{e}_1$ and $\mathbf{e}_2$ are disjunct and $p = \frac{2^{S_1}}{2^{S_1} + 2^{S_2}}$.*

Proof. If p is fixed, the upper variability bound of entropy is saturated only if $\mathbf{e}_1$ and $\mathbf{e}_2$ are disjunct by definition. The entropy maximum for the mixed ensemble can only be achieved when the elements are disjunct, for some value of p .

Now fix S_1 and S_2 . The entropy of the mixture depends only on p , so we need to find the p that maximizes the expression.

$$\begin{aligned}
0 &= \frac{dS}{dp} = \frac{d}{dp} S(\mathbf{e}) = \frac{d}{dp} (-p \log p - \bar{p} \log \bar{p} + pS_1 + \bar{p}S_2) \\
&= -\log p - 1 + \log \bar{p} + 1 + S_1 - S_2 \\
\log \frac{p}{\bar{p}} &= \log 2^{S_1} - \log 2^{S_2} \\
\log \frac{p}{1-p} &= \log \frac{2^{S_1}}{2^{S_2}} \\
p2^{S_2} &= (1-p)2^{S_1} \\
p(2^{S_1} + 2^{S_2}) &= 2^{S_1} \\
p &= \frac{2^{S_1}}{2^{S_1} + 2^{S_2}}
\end{aligned} \tag{1.24}$$

Having found the value of p that maximizes the entropy, we can calculate the maximum entropy.

$$\begin{aligned}
\bar{p} &= 1 - \frac{2^{S_1}}{2^{S_1} + 2^{S_2}} = \frac{2^{S_2}}{2^{S_1} + 2^{S_2}} \\
S &= S(\mathbf{e}) = -p \log p - \bar{p} \log \bar{p} + p S_1 + \bar{p} S_2 \\
&= -\frac{2^{S_1}}{2^{S_1} + 2^{S_2}} \log \frac{2^{S_1}}{2^{S_1} + 2^{S_2}} - \frac{2^{S_2}}{2^{S_1} + 2^{S_2}} \log \frac{2^{S_2}}{2^{S_1} + 2^{S_2}} \\
&\quad + \frac{2^{S_1}}{2^{S_1} + 2^{S_2}} \log 2^{S_1} + \frac{2^{S_2}}{2^{S_1} + 2^{S_2}} \log 2^{S_2} \\
&= \frac{2^{S_1}}{2^{S_1} + 2^{S_2}} \log (2^{S_1} + 2^{S_2}) + \frac{2^{S_2}}{2^{S_1} + 2^{S_2}} \log (2^{S_1} + 2^{S_2}) \\
&= \frac{2^{S_1} + 2^{S_2}}{2^{S_1} + 2^{S_2}} \log (2^{S_1} + 2^{S_2}) \\
\log 2^S &= \log (2^{S_1} + 2^{S_2}) \\
2^S &= 2^{S_1} + 2^{S_2}
\end{aligned} \tag{1.25}$$

Therefore the maximum entropy obtainable through a mixture is $S = \log(2^{S_1} + 2^{S_2})$ which is obtained when $\mathbf{e}_1$ and $\mathbf{e}_2$ are disjunct and $p = \frac{2^{S_1}}{2^{S_1} + 2^{S_2}}$. $\square$

Definition 1.26. Let $U \subseteq \mathcal{E}$ be the subset of an ensemble space. The **convex hull** of U , noted $\text{hull}(U)$ is the set of all possible mixtures that can be constructed with elements contained in U .

Corollary 1.27. The convex hull has the following properties

1. $U \subseteq \text{hull}(U)$
2. $U \subseteq V \implies \text{hull}(U) \subseteq \text{hull}(V)$
3. $\text{hull}(\text{hull}(U)) = \text{hull}(U)$

and is therefore a closure operations

Proof. 1. Every element of U is trivially a mixture of elements of U . Therefore $U \subseteq \text{hull}(U)$.

2. Let $\mathbf{e} \in \text{hull}(U)$. Then it is a mixture of some elements of U . Since $U \subseteq V$, then $\mathbf{e}$ is also the mixture of some elements of V and therefore $\mathbf{e} \in \text{hull}(V)$.

3. Since mixing is associative and commutative, a mixture of mixtures of U can be reexpressed as a mixture of U . Therefore for all $\mathbf{e} \in \text{hull}(\text{hull}(U))$ we have $\mathbf{e} \in \text{hull}(U)$. $\square$

Definition 1.28. Let $U \subseteq \mathcal{E}$ be the subset of an ensemble space. The **dimension** of U is defined as $\dim(U) = \sup(2^{S(\text{hull}(U))})$ if $U \neq \emptyset$ and $\dim(U) = 0$ otherwise.

Proposition 1.29. The dimension is a set function that is

1. non negative: $\dim(U) \in [0, +\infty]$
2. monotone: $U \subseteq V \implies \dim(U) \leq \dim(V)$
3. subadditive: $\dim(U \cup V) \leq \dim(U) + \dim(V)$
4. additive over disjunct sets: $U \perp V \implies \dim(U \cup V) = \dim(U) + \dim(V)$

Proof. 1. The dimension takes a subset of $\mathcal{E}$ and returns a real value and is therefore a set function. The exponential can only return non-negative values, therefore the dimension of a set is non-negative.

2. Let $U, V \subseteq \mathcal{E}$ such that $U \subseteq V$. If $U = \emptyset$, we have $\dim(U) = 0$. Since $\dim(V)$ is non-negative, $\dim(U) \leq \dim(V)$. If $U \neq \emptyset$, $\text{hull}(U) \subseteq \text{hull}(V)$ and therefore $2^{S(\text{hull}(U))} \subseteq 2^{S(\text{hull}(V))}$. This means that the supremum of the first set cannot be greater than the supremum of the second set, and therefore $\dim(U) \leq \dim(V)$. The dimension is a monotone set function.

3. Let $U, V \subseteq \mathcal{E}$ and let $\mathbf{e} \in \text{hull}(U \cup V)$. Then we can write $\mathbf{e} = p\mathbf{u} + \bar{p}\mathbf{v}$ for some $p \in [0, 1]$, $\mathbf{u} \in U$ and $\mathbf{v} \in V$. By 1.23 and the definition of dimension, $2^{S(\mathbf{e})} = 2^{S(\mathbf{u})} + 2^{S(\mathbf{v})} \leq \dim(U) + \dim(V)$. Since this is true for any element of $\text{hull}(U \cup V)$, the supremum of the exponential entropy cannot exceed the sum of the dimensions. Therefore $\dim(U \cup V) \leq \dim(U) + \dim(V)$, the dimension is subadditive.

4. Let $U, V \subseteq \mathcal{E}$ be two disjunct subsets. Let $\{\mathbf{u}_i\} \subset U$ be a sequence of ensembles such that $2^{S(\mathbf{u}_i)} \rightarrow \dim(U)$ and let $\{\mathbf{v}_i\} \subset V$ be a sequence of ensembles such that $2^{S(\mathbf{v}_i)} \rightarrow \dim(V)$. Consider $\mathbf{e}_i = p_i\mathbf{u}_i + \bar{p}_i\mathbf{v}_i$ where $p_i = \frac{2^{S(\mathbf{u}_i)}}{2^{S(\mathbf{u}_i)} + 2^{S(\mathbf{v}_i)}}$. Then, by 1.23, $2^{S(\mathbf{e}_i)} = 2^{S(\mathbf{u}_i)} + 2^{S(\mathbf{v}_i)}$. This means that $2^{S(\mathbf{e}_i)} \rightarrow \dim(U) + \dim(V)$. Therefore $\dim(U \cup V) \geq \dim(U) + \dim(V)$. Combining with the previous result, $\dim(U \cup V) = \dim(U) + \dim(V)$. Therefore dimension, as set function, is additive over disjunct sets of ensembles. $\square$

Definition 1.30. Let $\mathbf{e} \in \mathcal{E}$. The **probability measure** for $\mathbf{e}$ is defined as $p_{\mathbf{e}}(U) = \sup(\{p \in [0, 1] \mid \exists \mathbf{e}_1 \in \text{hull}(U), \mathbf{e}_2 \in \mathcal{E} \text{ s.t. } \mathbf{e} = p\mathbf{e}_1 + \bar{p}\mathbf{e}_2\})$ if $U \neq \emptyset$ and $p_{\mathbf{e}}(U) = 0$ otherwise.

Proposition 1.31. The probability measure for an ensemble is

1. non negative and unit bounded: $p_{\mathbf{e}}(U) \in [0, 1]$
2. monotone: $U \subseteq V \implies p_{\mathbf{e}}(U) \leq p_{\mathbf{e}}(V)$
3. subadditive: $p_{\mathbf{e}}(U \cup V) \leq p_{\mathbf{e}}(U) + p_{\mathbf{e}}(V)$

Proof. 1. The probability measure takes a subset of $\mathcal{E}$ and returns a real value and is therefore a set function. The coefficients used in mixtures are real values between zero and one. The supremum of a set of numbers between zero and one is between zero and one, and therefore the probability measure for an ensembles is non negative and unit bounded..

2. Let $U, V \subseteq \mathcal{E}$ such that $U \subseteq V$. If $U = \emptyset$, we have $p_{\mathbf{e}}(U) = 0$. Since $p_{\mathbf{e}}(V)$ is non-negative, $p_{\mathbf{e}}(U) \leq p_{\mathbf{e}}(V)$. If $U \neq \emptyset$, $\text{hull}(U) \subseteq \text{hull}(V)$ and therefore $\{p \in [0, 1] \mid \exists \mathbf{e}_1 \in \text{hull}(U), \mathbf{e}_2 \in \mathcal{E} \text{ s.t. } \mathbf{e} = p\mathbf{e}_1 + \bar{p}\mathbf{e}_2\} \subseteq \{p \in [0, 1] \mid \exists \mathbf{e}_1 \in \text{hull}(V), \mathbf{e}_2 \in \mathcal{E} \text{ s.t. } \mathbf{e} = p\mathbf{e}_1 + \bar{p}\mathbf{e}_2\}$. This means that the supremum of the first set cannot be greater than the supremum of the second set, and therefore $p_{\mathbf{e}}(U) \leq p_{\mathbf{e}}(V)$. The probability measure is a monotone set function.

3. Let $U, V \subseteq \mathcal{E}$ and let $p \in [0, 1]$ such that $\mathbf{e} = p\mathbf{e}_1 + \bar{p}\mathbf{e}_2$ for some $\mathbf{e}_1 \in \text{hull}(U \cup V)$ and $\mathbf{e}_2 \in \mathcal{E}$. Since $\mathbf{e}_1 \in \text{hull}(U \cup V)$, we can write $\mathbf{e}_1 = \lambda\mathbf{u} + \bar{\lambda}\mathbf{v}$ for some $\lambda \in [0, 1]$, $\mathbf{u} \in U$ and $\mathbf{v} \in V$. Therefore we have $\mathbf{e} = p\lambda\mathbf{u} + p\bar{\lambda}\mathbf{v} + \bar{p}\mathbf{e}_2$. By the definition of probability measure, we must have $p\lambda \leq p_{\mathbf{e}}(U)$ and $p\bar{\lambda} \leq p_{\mathbf{e}}(V)$, therefore $p = p\lambda + p\bar{\lambda} \leq p_{\mathbf{e}}(U) + p_{\mathbf{e}}(V)$. Since $p_{\mathbf{e}}(U \cup V)$ is the supremum for a set of p s for which the expression always holds, we have $p_{\mathbf{e}}(U \cup V) \leq p_{\mathbf{e}}(U) + p_{\mathbf{e}}(V)$. The probability measure is subadditive. $\square$

1.5 Entropic geometry

We now show that entropy imposes a geometric structure on the ensemble space.

Definition 1.32. Given two ensembles $\mathbf{e}_1, \mathbf{e}_2 \in \mathcal{E}$, the **mixing entropy**, also called *Jensen-Shannon divergence*, is the increase in entropy associated to their mixture. That is:

$$MS(\mathbf{e}_1, \mathbf{e}_2) = S\left(\frac{1}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2\right) - \left(\frac{1}{2}S(\mathbf{e}_1) + \frac{1}{2}S(\mathbf{e}_2)\right).$$

Corollary 1.33. The mixing entropy obeys the following bounds

$$0 \leq MS(\mathbf{e}_1, \mathbf{e}_2) \leq 1.$$

The lower bound is satisfied if and only if $\mathbf{e}_1 = \mathbf{e}_2$ and the upper bound is satisfied if and only if $\mathbf{e}_1 \perp \mathbf{e}_2$.

Proof. The bounds descend directly from the bounds on entropy. By strict concavity, $S(p_1\mathbf{e}_1 + p_2\mathbf{e}_2) \geq p_1S(\mathbf{e}_1) + p_2S(\mathbf{e}_2)$, which means $S(p_1\mathbf{e}_1 + p_2\mathbf{e}_2) - p_1S(\mathbf{e}_1) - p_2S(\mathbf{e}_2) \geq 0$, and in particular $S(\frac{1}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2) - \frac{1}{2}S(\mathbf{e}_1) - \frac{1}{2}S(\mathbf{e}_2) = MS(\mathbf{e}_1, \mathbf{e}_2) \geq 0$. By strict concavity, the equality holds if and only if $\mathbf{e}_1 = \mathbf{e}_2$.

By the upper variability bound, $S(p_1\mathbf{e}_1 + p_2\mathbf{e}_2) \leq I(p_1, p_2) + p_1S(\mathbf{e}_1) + p_2S(\mathbf{e}_2)$, which means $S(p_1\mathbf{e}_1 + p_2\mathbf{e}_2) - p_1S(\mathbf{e}_1) - p_2S(\mathbf{e}_2) \leq I(p_1, p_2)$, and in particular $S(\frac{1}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2) - \frac{1}{2}S(\mathbf{e}_1) - \frac{1}{2}S(\mathbf{e}_2) = MS(\mathbf{e}_1, \mathbf{e}_2) \leq I(\frac{1}{2}, \frac{1}{2}) = 1$. Given that $\mathbf{e}_1$ and $\mathbf{e}_2$ are disjunct by definition if they saturate the upper variability bound, the equality holds if and only if $\mathbf{e}_1$ and $\mathbf{e}_2$ are disjunct. $\square$

Proposition 1.34. In discrete and continuous classical cases, the mixing entropy coincides with the Jensen-Shannon divergence. In quantum spaces it coincides with the quantum Jensen-Shannon divergence.

Proof. Looking at the definitions, for example at https://en.wikipedia.org/wiki/Jensen%E2%80%93Shannon_divergence, one can see that

$$JSD(\mathbf{e}_1, \mathbf{e}_2) = S\left(\frac{1}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2\right) - \frac{1}{2}(S(\mathbf{e}_1) + S(\mathbf{e}_2)) = MS(\mathbf{e}_1, \mathbf{e}_2)$$

. The same is true for the quantum case:

$$QJSD(\mathbf{e}_1, \mathbf{e}_2) = S\left(\frac{1}{2}\mathbf{e}_1 + \frac{1}{2}\mathbf{e}_2\right) - \frac{1}{2}(S(\mathbf{e}_1) + S(\mathbf{e}_2)) = MS(\mathbf{e}_1, \mathbf{e}_2)$$

.

$\square$

Definition 1.35. An ensemble space is **geometric** if it is complemented and has a twice differentiable entropy with respect to the mixing coefficients.

Corollary 1.36. A geometric ensemble space is a convex subset of a smooth vector space.

Proof. Since a geometric ensemble space is complemented, it embeds into a real vector space. The differentiable structure is imposed by requiring that any parameter used to define families of ensembles must be twice differentiable with respect to the entropy. $\square$

Remark. These requirements are useful to leverage all definitions of differential geometry, which require a local vector space structure and differentiability. It may be the case that convex structure of the ensemble space is enough, and that the differentiability of the entropy can be derived by showing the continuity of the mixing entropy.

For now, we are going to assume the above. This is already a generalization as we work with a generic vector space and not a manifold. The hope is that this can be a rigorous foundation for classical and quantum field theories.

Definition 1.37. Given a geometric ensemble $\mathbf{e} \in \mathcal{E}$, the **tangent space** at $\mathbf{e}$, noted $T_{\mathbf{e}}$, is the space of all infinitesimal variations $\delta\mathbf{e}$. The **norm** of a variation $\delta\mathbf{e} \in T_{\mathbf{e}}$ is given by

$$\|\delta\mathbf{e}\|_{\mathbf{e}} = \sqrt{MS(\mathbf{e}, \mathbf{e} + \delta\mathbf{e})}.$$

The **metric tensor** (i.e. the inner product between two variations $\delta\mathbf{e}_1, \delta\mathbf{e}_2 \in T_{\mathbf{e}}$) is given by

$$g_{\mathbf{e}}(\delta\mathbf{e}_1, \delta\mathbf{e}_2) = \frac{1}{2} \left(\|\delta\mathbf{e}_1 + \delta\mathbf{e}_2\|_{\mathbf{e}}^2 - \|\delta\mathbf{e}_1\|_{\mathbf{e}}^2 - \|\delta\mathbf{e}_2\|_{\mathbf{e}}^2 \right).$$

Theorem 1.38. Let $\mathcal{E}$ be a geometric ensemble space. Then

$$\|\delta\mathbf{e}\|_{\mathbf{e}}^2 = -\frac{1}{8} \frac{\partial^2 S}{\partial \mathbf{e}^2}(\delta\mathbf{e}, \delta\mathbf{e})$$

and

$$g_{\mathbf{e}}(\delta\mathbf{e}_1, \delta\mathbf{e}_2) = -\frac{1}{8} \frac{\partial^2 S}{\partial \mathbf{e}^2}(\delta\mathbf{e}_1, \delta\mathbf{e}_2).$$

Let $V \subseteq \mathcal{E}$ be a finite dimensional subregion. Then V is a Riemannian manifold with $g_{\mathbf{e}}$ as the metric tensor and $\|\cdot\|_{\mathbf{e}}$ as the norm.

Proof. To recover the first two expression, we simply have to calculate the leading term. Since the entropy is twice differentiable, we can expand it as

$$S(\mathbf{e} + \delta\mathbf{e}) = S(\mathbf{e}) + \frac{\partial S}{\partial \mathbf{e}}\delta\mathbf{e} + \frac{1}{2} \frac{\partial^2 S}{\partial \mathbf{e}^2}\delta\mathbf{e}\delta\mathbf{e} + O(\delta\mathbf{e}^3). \quad (1.39)$$

Expanding the definition of MS , we have

$$\begin{aligned} MS(\mathbf{e}, \mathbf{e} + \delta\mathbf{e}) &= S\left(\frac{1}{2}\mathbf{e} + \frac{1}{2}(\mathbf{e} + \delta\mathbf{e})\right) - \frac{1}{2}S(\mathbf{e}) - \frac{1}{2}S(\mathbf{e} + \delta\mathbf{e}) \\ &= S\left(\mathbf{e} + \frac{1}{2}\delta\mathbf{e}\right) - \frac{1}{2}S(\mathbf{e}) - \frac{1}{2}S(\mathbf{e} + \delta\mathbf{e}) \\ &= S(\mathbf{e}) + \frac{\partial S}{\partial \mathbf{e}}\frac{1}{2}\delta\mathbf{e} + \frac{1}{2}\frac{\partial^2 S}{\partial \mathbf{e}^2}\frac{1}{2}\delta\mathbf{e}\frac{1}{2}\delta\mathbf{e} + O(\delta\mathbf{e}^3) \\ &\quad - \frac{1}{2}S(\mathbf{e}) - \frac{1}{2}\left(S(\mathbf{e}) + \frac{\partial S}{\partial \mathbf{e}}\delta\mathbf{e} + \frac{1}{2}\frac{\partial^2 S}{\partial \mathbf{e}^2}\delta\mathbf{e}\delta\mathbf{e} + O(\delta\mathbf{e}^3)\right) \\ &= S(\mathbf{e}) + \frac{1}{2}\frac{\partial S}{\partial \mathbf{e}}\delta\mathbf{e} + \frac{1}{8}\frac{\partial^2 S}{\partial \mathbf{e}^2}\delta\mathbf{e}\delta\mathbf{e} \\ &\quad - S(\mathbf{e}) - \frac{1}{2}\frac{\partial S}{\partial \mathbf{e}}\delta\mathbf{e} - \frac{1}{4}\frac{\partial^2 S}{\partial \mathbf{e}^2}\delta\mathbf{e}\delta\mathbf{e} + O(\delta\mathbf{e}^3) \\ &= -\frac{1}{8}\frac{\partial^2 S}{\partial \mathbf{e}^2}\delta\mathbf{e}\delta\mathbf{e} + O(\delta\mathbf{e}^3). \end{aligned} \quad (1.40)$$

Therefore

$$\|\delta\mathbf{e}\|^2 = MS(\mathbf{e}, \mathbf{e} + \delta\mathbf{e}) = -\frac{1}{8}\frac{\partial^2 S}{\partial \mathbf{e}^2}(\delta\mathbf{e}, \delta\mathbf{e}).$$

We can now substitute the norm to the definition of metric tensor. We have

$$\begin{aligned}
g_e(\delta e_1, \delta e_2) &= \frac{1}{2} (\|\delta e_1 + \delta e_2\|^2 - \|\delta e_1\|^2 - \|\delta e_2\|^2) \\
&= \frac{1}{2} \left(-\frac{1}{8} \frac{\partial^2 S}{\partial e^2}(\delta e_1 + \delta e_2, \delta e_1 + \delta e_2) + \frac{1}{8} \frac{\partial^2 S}{\partial e^2}(\delta e_1, \delta e_1) + \frac{1}{8} \frac{\partial^2 S}{\partial e^2}(\delta e_2, \delta e_2) \right) \\
&= -\frac{1}{16} \left(\frac{\partial^2 S}{\partial e^2}(\delta e_1, \delta e_1) + \frac{\partial^2 S}{\partial e^2}(\delta e_1, \delta e_2) + \frac{\partial^2 S}{\partial e^2}(\delta e_2, \delta e_1) + \frac{\partial^2 S}{\partial e^2}(\delta e_2, \delta e_2) \right) \\
&\quad - \frac{\partial^2 S}{\partial e^2}(\delta e_1, \delta e_1) - \frac{\partial^2 S}{\partial e^2}(\delta e_2, \delta e_2) \\
&= -\frac{1}{8} \frac{\partial^2 S}{\partial e^2}(\delta e_1, \delta e_2)
\end{aligned} \tag{1.41}$$

Let us now take a finite dimensional region $V \subseteq \mathcal{E}$. Since V is a finite dimensional subregion of a real vector space, it will be a manifold. It will have a smooth structure inherited from the ensemble space.

Let us now show that g is a metric tensor. Restricted to V , $g|_V$ is the Hessian of the entropy S . Therefore it is a symmetric rank two tensor which means a symmetric bilinear function of the arguments. Since the entropy is strictly concave, the Hessian is negative definite and therefore g is positive definite. Therefore $(V, g|_V)$ is a Riemannian manifold. $\square$

Remark. We have shown that, under very general conditions, a generalized ensemble space is a geometric space, where the notion of distance and angle are fully defined by the entropy. That is, all geometric structures in physics are entropic structures.

1.6 Ensemble subspaces

We now want to define a notion of subspace in a way that recovers the notions of subspaces we have in both classical and quantum mechanics. Given that orthogonality in all three cases corresponds to ensembles being disjoint, disjoint ensembles will need to belong to different subspaces. The idea is to reconstruct subspaces only from the disjoint relationship.

Irreflexive symmetric relations and topped $\cap$ -structures

We will first prove a more general set of results, on a generic set (not necessarily an ensemble space) with an irreflexive symmetric relation (not necessarily disjointness of ensembles). This will give us a more general tool to investigate possible different notion of subspaces.

Definition 1.42. Let X be a set and $R \subseteq X \times X$ a symmetric relation. Given a subset $U \subseteq X$, we define the **R -complement** to be

$$U^{RC} = \{a \in X \mid \forall b \in U, aRb\}.$$

Proposition 1.43. Let X be a set and $R \subseteq X \times X$ a symmetric relation. Then:

1. $U \subseteq V \implies V^{RC} \subseteq U^{RC}$
2. $U \subseteq (U^{RC})^{RC}$
3. $U^{RC} = ((U^{RC})^{RC})^{RC}$
4. $U^{RC} = (V^{RC})^{RC} \iff (U^{RC})^{RC} = V^{RC}$

5. $(\bigcup_{i \in I} U_i)^{RC} = \bigcap_{i \in I} (U_i)^{RC}$
6. $\emptyset^{RC} = X$

Proof. 1. Suppose $a \in V^{RC}$. Then, by definition, $\forall b \in V, aRb$. Since $U \subset V$, it is also true that $\forall b \in U, aRb$. Therefore $a \in U^{RC}$ by definition. Since a was arbitrary, $V^{RC} \subseteq U^{RC}$.

2. By expanding the definition of complement, we have $(U^{RC})^{RC} = \{a \in X \mid \forall b \in U^{RC}, aRb\} = \{a \in X \mid \forall b \in \{c \in X \mid \forall d \in U, cRd\}, aRb\} = \{a \in X \mid \forall b \in X \text{ s.t. } (\forall d \in U, bRd), aRb\}$.

Let $a \in U$ and let $b \in X$ such that $\forall d \in U, bRd$. Since $a \in U$ and bRd for all $b \in U$, we have bRa in particular. Since R is symmetric, aRb . Given that b was arbitrary, we conclude that $\forall b \in X \text{ s.t. } (\forall d \in U, bRd), aRb$. Therefore $a \in (U^{RC})^{RC}$ by definition of complement. Given that a was arbitrary, $U \in (U^{RC})^{RC}$.

3. We again expand the definition and have $((U^{RC})^{RC})^{RC} = \{a \in X \mid \forall b \in (U^{RC})^{RC}, aRb\}$.

Let $x \in ((U^{RC})^{RC})^{RC}$. Then $\forall b \in (U^{RC})^{RC}, xRb$ by definition of the complement. Since by 1. $U \subset (U^{RC})^{RC}$, we can restrict the previous expression to only the elements of U , and therefore $\forall b \in U, xRb$. But this means that $x \in U^{RC}$ by definition of the complement. Since x was arbitrary, $((U^{RC})^{RC})^{RC} \subseteq U^{RC}$. But by 1., we also have $U^{RC} \subseteq ((U^{RC})^{RC})^{RC}$ since U^{RC} is just a set onto which we can apply the complement twice. By two-way containment, we have $U^{RC} = ((U^{RC})^{RC})^{RC}$.

4. Let $U, V \subseteq X$ such that $U^{RC} = (V^{RC})^{RC}$. Applying the complement on each side, $(U^{RC})^{RC} = ((V^{RC})^{RC})^{RC}$. By the previous property $((V^{RC})^{RC})^{RC} = V^{RC}$ and therefore $(U^{RC})^{RC} = V^{RC}$. Switching U and V proves the other direction.

5. We have:

$$\begin{aligned}
 \left(\bigcup_{i \in I} U_i\right)^{RC} &= \{a \in X \mid \forall b \in \bigcup_{i \in I} U_i, aRb\} \\
 &= \{a \in X \mid \forall U_i, \forall b \in U_i, aRb\} \\
 &= \{a \in X \mid \forall U_i, a \in \{c \in X \mid \forall b \in U_i, cRb\}\} \\
 &= \bigcap_{i \in I} \{c \in X \mid \forall b \in U_i, cRb\} \\
 &= \bigcap_{i \in I} (U_i)^{RC}
 \end{aligned}$$

Note that this property does not rely on the symmetry of R .

6. Let $a \in X$. There is no $b \in \emptyset$ such that aRb . Therefore $\forall b \in \emptyset, aRb$. This means that $a \in \emptyset^{RC}$. Since a was arbitrary, $X = \emptyset^{RC}$. $\square$

Proposition 1.44. *Let X be a set and $R \subseteq X \times X$ a symmetric and irreflexive relation. Then*

1. $U \cap U^{RC} = \emptyset$
2. $X^{RC} = \emptyset$

Proof. 1. Let $a \in U$. Since R is irreflexive, aRa is false. Therefore it is not true that, for all $b \in U, aRb$. This means that $a \notin U^{RC}$. Since that a was arbitrary, $U \cap U^{RC} = \emptyset$.

2. Suppose $a \in X^{RC}$. Then for all $b \in X, aRb$. In particular, we would have aRa , which can't be true since R is irreflexive. Therefore $a \notin X^{RC}$ and, since a is arbitrary, $X^{RC} = \emptyset$. $\square$

Definition 1.45. *Let X be a set and $R \subseteq X \times X$ a symmetric relation. Let $U \subseteq X$. The **R -subspace generated by U** is $\langle U \rangle_R = (U^{RC})^{RC}$. An **R -subspace** of X is a set $U \subseteq X$ such*

that $U = \langle U \rangle_R$. The **lattice of R -subspaces** is the set $\mathfrak{L} = \{U \subseteq X \mid U = \langle U \rangle_R\}$ ordered by inclusion.

Corollary 1.46. *The lattice of R -subspaces $\mathfrak{L}$ is a topped $\cap$ -structure on X and therefore is also a complete lattice.*

Proof. The set $\mathfrak{L}$ is a collection of subsets of X . Let $\{U_i\}_{i \in I} \subseteq \mathfrak{L}$ be a non-empty family. Then, using the definition of subspace, and the third and fifth properties of 1.43, we have

$$\begin{aligned} \bigcap_{i \in I} U_i &= \bigcap_{i \in I} (U_i^{RC})^{RC} = (\bigcup_{i \in I} U_i^{RC})^{RC} \\ &= (((\bigcup_{i \in I} (U_i^{RC})^{RC})^{RC})^{RC} = ((\bigcap_{i \in I} (U_i^{RC})^{RC})^{RC})^{RC} \\ &= ((\bigcap_{i \in I} U_i)^{RC})^{RC}. \end{aligned}$$

Therefore $\bigcap_{i \in I} U_i \in \mathfrak{L}$. This means $\mathfrak{L}$ is an $\cap$ -structure. Using the second property of 1.44, $X = \emptyset^{RC} = (X^{RC})^{RC}$. Therefore $\mathfrak{L}$ is a topped $\cap$ -structure. This also means that it is a complete lattice. $\square$

Corollary 1.47. *The R -subspace operator satisfies the following properties*

1. $U \subseteq \langle U \rangle_R$
2. $U \subseteq V \implies \langle U \rangle_R \subseteq \langle V \rangle_R$
3. $\langle \langle U \rangle_R \rangle_R = \langle U \rangle_R$

and is therefore a closure operation.

Proof. 1. The first property is true by the second property of 1.44.

2. Using the first property of 1.44 we have $U \subseteq V$ implies $V^{RC} \subseteq U^{RC}$ which in turns implies $(U^{RC})^{RC} \subseteq (V^{RC})^{RC}$. Therefore $\langle U \rangle_R \subseteq \langle V \rangle_R$.

3. Using the fifth property of 1.44 we have $\langle \langle U \rangle_R \rangle_R = (((U^{RC})^{RC})^{RC})^{RC} = (U^{RC})^{RC} = \langle U \rangle_R$. $\square$

Proposition 1.48. *Let X be a set and $R \subseteq X \times X$ a symmetric and irreflexive relation. Then*

1. $\langle U \rangle_R$ is the smallest R -subspace containing U
2. if $U, V \in \mathfrak{L}$ then $U = V^{RC} \iff U^{RC} = V$

Proof. 1. Let $U \subseteq X$ and $V \in \mathfrak{L}$ such that $U \subseteq V$ and $V \subseteq \langle U \rangle_R$. Since $U \subseteq V$, using the second property of 1.47, $\langle U \rangle_R \subseteq \langle V \rangle_R = V$. Since $V \subseteq \langle U \rangle_R$ and $\langle U \rangle_R \subseteq V$, $\langle U \rangle_R = V$. This means that no R -subspace that contains U is smaller than $\langle U \rangle_R$.

2. Since $U, V \in \mathfrak{L}$, $U = (U^{RC})^{RC} = V^{RC}$ which, by the fourth property of 1.43, implies $U^{RC} = (V^{RC})^{RC} = V$. Switching U and V proves the other direction. $\square$

Proposition 1.49. *Let X be an inner product space and $R = \{(a, b) \mid \langle a, b \rangle = 0\}$. Then:*

1. R is an irreflexive and symmetric relation
2. $U^{RC} = U^\perp$
3. $\langle U \rangle_R = \text{cl}_X(\text{span}(U))$

Proof. 1. Given that the inner product is symmetric, so will be R . Given that no vector is orthogonal to itself, R is irreflexive.

2. The orthogonal complex is defined as $U^\perp = \{a \in X \mid \forall b \in U, \langle a, b \rangle = 0\}$. Since $aRb \iff \langle a, b \rangle = 0$, $U^{RC} = U^\perp$.

3. We have $\langle U \rangle_R = (U^\perp)^\perp$ which returns the smallest closed subspace that contains U . $\square$

This shows that full structure of subspaces of an inner product space can be fully recovered only from pairwise orthogonality.

Subspaces of ensemble spaces and states

We will now use the previous results where X is an ensemble space and R is the disjunct relation between two ensembles.

Proposition 1.50. *Let $\mathcal{E}$ be an ensemble space, disjunctness $\perp$ is an irreflexive symmetric relation.*

Proof. Since any ensemble is not distinct from itself, it is neither disjunct from itself. Therefore disjunctness is irreflexive. Two elements are disjunct if they saturate the upper entropy bound. This does not depend on their order. Therefore disjunctness is symmetric. $\square$

Definition 1.51. *Let $\mathcal{E}$ be an ensemble space and $X \subseteq \mathcal{E}$ be a subset. The **disjunct complement** $X^\perp \subseteq \mathcal{E}$ is the set of all ensembles that are disjunct from all elements of X . An **ensemble subspace** is a subset $X \in \mathcal{E}$ such that $X = (X^\perp)^\perp$.*

Remark. Since subspaces are subsets of the ensembles space, the dimension and probability measure are automatically defined on subspaces.

Conjecture 1.52. *Let $\mathcal{E}$ be a discrete classical ensemble space. Then*

1. *the lattice of subspaces $\mathfrak{L}$ is the set of all possible simplexes generated by all possible combinations of extreme points*
2. *the dimension of each subspace is the number of extreme points*

Proof. Let $\mathcal{E}$ be a discrete classical ensemble space. Then each ensemble is a sequence p_i such that $\sum_i p_i = 1$. The dimensionality of the space fixes the range of i . Note that, given that the space is discrete, the collection p_i can be understood as a continuous function of the pure elements.

Note that two probability distributions will be disjunct if and only if their supports are disjoint. Therefore disjunctness for a discrete classical ensemble space is the same as having disjoint support. This means that a subspace is given by all possible probability distributions over a subset of all pure states.

If U is the set of all probability distributions over n cases, the distribution with highest entropy is achieved by a continuous distribution and will be $S(p_i) = \log n$. Therefore the dimension of U will be n . $\square$

Conjecture 1.53. *Let $\mathcal{E}$ be a continuous classical ensemble space. Then*

1. *the lattice of subspaces $\mathfrak{L}$ is the set of all subspaces of continuous probability densities with support within an open set that is the interior of its own closure*
2. *the dimension of each subspace is the Liouville measure*

Proof. Let $\mathcal{E}$ be a continuous classical ensemble space. Then each ensemble is a continuous function for a symplectic manifold X that integrates to one. Since the function is continuous, its support must be an open set.

Consider two continuous functions that integrate to one. They will be disjunct if and only if their supports are disjoint. Therefore disjunctness for a continuous classical ensemble space is the same as having disjoint support.

Take a set of ensembles $U \subseteq \mathcal{E}$. An ensemble will be disjunct from all ensembles in U if and only if its support is disjoint from the union of all the supports of all the elements of U . Therefore $U^\perp$ is the set of all ensembles whose support is a subset of the exterior of the union of all supports of elements of U . With a similar logic, $(U^\perp)^\perp$ is the set of all ensembles whose support is a subset of the exterior of the support of elements of $U^\perp$. This corresponds to the interior of the closure of the union of all supports of elements of U . Therefore, any subspace is a set of all continuous probability densities with support within an open set $V \subset X$ that is the interior of its own closure.

Now take an open set $V \subseteq X$ that is the interior of its own closure and consider the space U of all continuous probability densities that have support within V . Then $U^\perp$ is the subspace of all continuous probability densities that have support within $\text{ext}(V)$ and $(U^\perp)^\perp$ is the subspace of all continuous probability densities that have support within $\text{ext}(\text{ext}(V)) = \text{int}(\overline{V}) = V$. Therefore U is a subspace.

Let $U \in \mathcal{L}$ be an ensemble subspace. If $e \in U$, then $S(e) \leq \dim(U)$. Moreover, there must be a sequence e_i such that $S(e_i) \rightarrow \dim(U)$. Let $V \subseteq X$ be the open set corresponding to the largest support of all elements in U . Then $S(e) \leq \log \mu(V)$ as the highest entropy is given by a uniform distribution over V . The uniform distribution is discontinuous, and therefore is not a member of $\mathcal{E}$. However, we can construct a sequence of continuous functions whose limit is the uniform distribution. This would correspond to a sequence of ensembles for which $S(e_i) \rightarrow \log \mu(V)$. Therefore $\log \mu(V)$ is the supremum of the entropy for the subspace, and therefore the dimension is $\mu(V)$, the Liouville measure. $\square$

Conjecture 1.54. *Let $\mathcal{E}$ be a quantum ensemble space. Then*

1. *the lattice of subspaces $\mathcal{L}$ is the set of all possible density operators for each subspace of the Hilbert space*
2. *the dimension of each subspace is the dimension of the corresponding Hilbert space*

Definition 1.55. *A **pure state** is a limit of an ensemble with the smallest entropy. Formally, a pure state $s \subset \mathcal{L}$ is a non-empty collection of subspaces such that*

1. *if $\{X_i\}_{i \in I} \in s$ then $\bigcap_{i \in I} X_i \in s$*
2. *if $X \in s$ and $Y \in \mathcal{L}$ such that $X \subseteq Y$, then $Y \in s$*
3. *if $X \in s$ and $Y \in \mathcal{L}$ such that $Y \subseteq X$ and $Y \neq \emptyset$, then there exists $Z \in s$ such that $Z \subset X$.*

The set of all pure states for an ensemble space is noted $S(\mathcal{E})$.

Definition 1.56. *The set of states of a subspace $X \in \mathcal{L}$ is the set $\text{supp}(X) = \{s \in S(\mathcal{E}) \mid X \in s\}$. The topology of the pure states is the topology over the set of pure states generated by the set of states of all subspaces.*

Definition 1.57.

Definition 1.58. Given an ensemble space $\mathcal{E}$, its **state space** is the set of all pure states $S(\mathcal{E})$ equipped with

1. the topology generated by the set of states of all subspaces
2. the measure

Probability context

Definition 1.59. A **probability context** is a set of subspaces $C \subseteq \mathfrak{L}$ such that

1. if $\{X_i\}_{i \in I} \in C$ then $\bigcap_{i \in I} X_i \in C$ and $\bigvee_{i \in I} X_i \in C$
2. if $X, Y \in C$, then either $X \subset Y$, $Y \subset X$ or $X \perp Y$
3. if $X, Y \in C$, then $X \vee Y \in C$

Conjecture 1.60. Let $\mathcal{E}$ be a discrete finite classical ensemble space. Then

1. the lattice of subspaces $\mathfrak{L}$ is the set of all possible simplexes generated by all possible combinations of extreme points
2. the dimension of each subspace is the number of extreme points
3. the pure states are the extreme points.

Conjecture 1.61. Let $\mathcal{E}$ be a continuous classical ensemble space. Then

1. the lattice of subspaces $\mathfrak{L}$ is the set of all subspaces of probability measures with support over an open set U
2. the dimension of each subspace is the Liouville measure
3. the pure states are the points of the symplectic manifold.

Conjecture 1.62. Let $\mathcal{E}$ be a quantum ensemble space. Then

1. the lattice of subspaces $\mathfrak{L}$ is the set of all possible density operators for each subspace of the Hilbert space
2. the dimension of each subspace is the dimension of the corresponding Hilbert space
3. the pure states are the subspaces of dimension one (i.e. the rays of the Hilbert space).

Conjecture 1.63. Let $\mathcal{E}$ be a reducible ensemble space and $X, Y \subseteq \mathcal{E}$ two subspaces. Then $\dim(X \vee Y) = \dim(X) + \dim(Y)$.

Definition 1.64. An ensemble is **decomposable** if it is the mixture of two or more components. An ensemble space is **infinitesimally decomposable** if all ensembles are decomposable. An ensemble space is **finitely decomposable** if any ensemble can be decomposed into elements that are not decomposable.

Proposition 1.65. The discrete classical ensemble space and the quantum ensemble space are finitely decomposable. The continuous classical space is infinitesimally decomposable.

Proof. Pure states in discrete classical and quantum ensembles can no longer be decomposed, and every other ensemble is the mixture of pure states. Therefore these spaces are finitely decomposable.

The ensembles of a continuous classical ensemble spaces are probability distributions that are continuous and therefore must have finite support. However, given any finite support, we

can always find distribution with smaller support and therefore decompose any distribution in the sum of two distribution. Therefore the space is infinitesimally decomposable. $\square$

Conjecture 1.66. *An ensemble space $\mathcal{E}$ is a finite discrete classical ensemble space if and only if the following properties are all satisfied*

1. $\mathcal{E}$ is reducible (i.e. classical)
2. every state with zero entropy is non-decomposable
3. the entropy is bounded

Proof. Let us first show that all finite discrete classical $\square$

Remark. The non-decomposable ensembles are exactly the extreme points of the convex space.

1.7 Lessons learned

Subspaces from convex structure

Insight 1.67. *Convex structure, by itself, cannot define subspaces.*

Initially, we tried recovering the notion of subspaces purely from the convex structure. We did make some progress in the finite dimensional case, but ultimately this does not work. The ultimate problem is that it cannot be generalized to the classical continuous case. The problem of rate of convergence is described below in the example. Even if that problem is fixed, there would be nothing to determine the dimensionality of U , and we could create convex maps that “stretch” the space.

The attempt was the following:

Definition 1.68. *Let $\mathcal{E}$ be a convex space and $X \subseteq \mathcal{E}$ be a subset. We say that X is a **subspace** of $\mathcal{E}$ if it contains all the convex combinations and all the components of its elements. That is, for every $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in \mathcal{E}$ and $\lambda \in (0, 1)$ such that $p\mathbf{e}_1 + \bar{p}\mathbf{e}_2 = \mathbf{e}_3$ we have:*

- $\mathbf{e}_1, \mathbf{e}_2 \in X$ implies $\mathbf{e}_3 \in X$
- $\mathbf{e}_3 \in X$ implies $\mathbf{e}_1, \mathbf{e}_2 \in X$

Definition 1.69. *Let $\mathcal{E}$ be a convex space and $X \subseteq \mathcal{E}$. The **convex span** of X , noted $\text{cospn}(X)$, is the smallest subspace containing X .*

Remark. As defined, the convex span of two elements will include all their possible mixtures (i.e. the segment that connects them), all possible decompositions (i.e. all lines that pass through them) plus, recursively, all other mixtures and decompositions that can be reached from those. Physically, the idea is that not all ensembles, pure states in particular, cannot be physically realized. Therefore, the inclusion of pure states in our convex space stems from a theoretical idealization useful to decompose and study the problem. It makes sense, then, that a subspace comes with all its idealizations.

Example 1.70. Classical discrete subspaces. Let S be a set of n possible discrete states and let $\mathcal{E}$ the space of probability distributions over the set S (i.e. $\mathcal{E}$ is an n -simplex and S are its extreme points). A subspace X of $\mathcal{E}$ is a convex hull of a subset U of S . That is, a subspace

of $\mathcal{E}$ is the space of probability distributions over a subset of the cases. Geometrically, it is one of the sides (possibly recursively) of the simplex.

To see this, first note that the convex hull X of any subset U of extreme points S is a subspace. In fact, it will contain all convex combinations of U , and any element can only be decomposed in convex combinations of U . Second, note that only convex hulls of a subset of extreme points can be a subspace. In fact, any element of $\mathcal{E}$ can be expressed as a non-trivial convex combination of a set of extreme points U . Therefore, if an element is present in a subspace X , then $U \subset X$, which means all elements of the convex hull of U are in X .

Remark. The definition does not entirely work in the continuous case. Let S be a symplectic manifold and let $\mathcal{E}$ be the space of probability distributions over S (i.e. the space of continuous functions over S that are integrable and integrate to one). We would like to have a definition for which a subspace of $\mathcal{E}$ is a set of functions whose support is a subset of an open region $U \subseteq S$. That is, a subspace contains all probability distributions defined over a subset of the cases. One problem is that continuity forces the function to go to zero on the boundary of U , and the speed of the convergence cannot be change with a finite convex combination. That is, the convex combination of functions that go down like an exponential will also go down like an exponential. The issue seems to be related to finding the correct closure for infinite convex combinations in the definition of convex space.

Example 1.71. Quantum subspaces. Let $\mathcal{H}$ be an n -dimensional Hilbert space and let $\mathcal{E}$ be the space of density matrices (i.e. positive semi-definite self-adjoint operators with trace one). A subspace X of $\mathcal{E}$ is the space of density matrices of a subspace U of $\mathcal{H}$. That is, a subspace of $\mathcal{E}$ is the space of mixed states over a subspace of pure states.

To see this, first note that the space of density matrices X of a subspace U of $\mathcal{H}$ is a subspace of $\mathcal{E}$. In fact, X it will contain all convex combinations of its elements. Moreover, any element $x \in X$ can only be decomposed in a convex combination of pure states of U . Therefore any convex decomposition of x has all its elements in X . Second, note that only the space of density matrices X of a subspace U of $\mathcal{H}$ is a subspace of $\mathcal{E}$. In fact, any element x of $\mathcal{E}$ can be expressed as a non-trivial convex combination of orthogonal pure states, its eigenstates. These elements will span a subspace U of $\mathcal{H}$. From those elements, we can construct an equal mixture which represents the maximally mixed states and, mathematically, is the identity operator I/m divided by the number of elements $m \leq n$ of U . The equal mixture of any orthogonal basis of U will also give the maximally mixed state. Therefore, given an element x , any subspace that contain x will also contain a basis of U , the maximally mixed state I/n , all possible basis of U , which means all the pure states, and finally all convex combinations of the pure states, which means all possible density matrices, all possible mixed states.

Definition 1.72. A convex space $\mathcal{E}$ is **closed** if it contains all its extreme points.

Definition 1.73. Let $\mathcal{E}$ be a convex space and $X \subset \mathcal{E}$ be a subspace of $\mathcal{E}$. The **dimension** of X , noted $\dim(X)$, is the minimum number of extreme points whose convex span is X .

Conjecture 1.74. Let $\mathcal{E}$ be closed convex space of finite dimensions. Then:

- $\mathcal{E}$ is a simplex if and only if every two dimensional subspace is a line segment
- the set of extreme points of $\mathcal{E}$ is a complex projective space if and only if every two dimensional subspace is a two dimensional sphere ($\mathcal{E}$ is the space of density matrices).

Remark. The only if is easy to show based on the discussion above. The if part does not actually work. Take 6 points in a 4 dimensional euclidean space such that they form two triangles that intersect at a single point. All two dimensional convex spaces will be segments. However, the whole convex hull is not a simplex.