

Contents

1	Geometry of probability distributions	3
1.1	Manifolds of probability distributions	3
1.1.1	Space of probability distributions	3
1.1.2	Statistical models and manifolds	4
1.1.3	The tangent space and its representations	5
1.2	The information metric	7
1.2.1	Relative entropy	7
1.2.2	Squared Riemannian distances	9
1.2.3	Divergences	11
1.2.4	The Fisher information metric	12
1.2.5	The geometry of \mathcal{P}	13
1.3	Parameter estimation	14
1.3.1	Unbiased estimators	14
1.3.2	Variance and expectation value	15
1.3.3	The Cramér-Rao bound	17
2	Geometry of quantum states	19
2.1	Quantum mechanics	19
2.1.1	The postulates of quantum mechanics	19
2.1.2	PVM and observables	22
2.1.3	Quantum states	24
2.1.4	Density operators	25
2.2	The manifold of quantum states	29
2.2.1	Geometry of the Hilbert space	29
2.2.2	Fiber bundle structure of the Hilbert space	30
2.2.3	The Fubini-Study metric	31
3	Quantum information	36
3.1	Quantum information geometry	36
3.1.1	Observables and the tangent space	36
3.1.2	Quantum Fisher information	38

3.1.3 Symmetric covariance	38
--------------------------------------	----

Chapter 1

Geometry of probability distributions

Quantum states, as we will see, can be thought of as a generalization of probability distributions. In this chapter, we will study probability distributions from a geometrical point of view, and in this framework, we shall prove the Cramer-Rao bound. We will only consider probability distributions defined on finite sample spaces since this is all we need for finite-dimensional pure quantum states.

1.1 Manifolds of probability distributions

1.1.1 Space of probability distributions

Consider a *random process* and the set \mathcal{X} of all its possible outcomes. We call this set the *sample space*, and we will only consider random processes for which it is finite. Then, a *probability distribution* on \mathcal{X} is a function $p \in \mathbb{R}^{\mathcal{X}} := \{f \mid f: \mathcal{X} \rightarrow \mathbb{R}\}$ which satisfies

$$p(x) \geq 0 \quad \forall x \in \mathcal{X} \quad \text{and} \quad \sum_{x \in \mathcal{X}} p(x) = 1 \tag{1.1}$$

where $p(x)$ represents the probability of the outcome x .

Further, every function $A \in \mathbb{R}^{\mathcal{X}}$ represents a real *random variable*, as it maps every outcome of a random process to a number. Then the *expectation value* of A when the underlying probability distribution is p is expressed by

$$\mathbb{E}_p[A] := \sum_{x \in \mathcal{X}} p(x)A(x) \tag{1.2}$$

Also, given two random variables A, B their *covariance* is

$$\text{Cov}_p[A, B] := \mathbb{E}_p[(A - \mathbb{E}_p[A])(B - \mathbb{E}_p[B])] \quad (1.3)$$

and so the *variance* of a random variable A is

$$\text{V}_p[A] := \text{Cov}_p[A, A] = \mathbb{E}_p[(A - \mathbb{E}_p[A])^2] \quad (1.4)$$

Let now N be the cardinality of \mathcal{X} . To have a picture of $\mathbb{R}^{\mathcal{X}}$ we can index the outcomes and consider the natural isomorphism between $\mathbb{R}^{\mathcal{X}}$ and \mathbb{R}^N

$$f \leftrightarrow (f(x_1), \dots, f(x_N)) \quad (1.5)$$

then it's easy to recognize that the *space of probability distributions* is a convex subset of the affine subspace $\mathcal{A}_1 := \{f \in \mathbb{R}^{\mathcal{X}} \mid \sum_{x \in \mathcal{X}} f(x) = 1\}$. In particular, it is the set resulting from the convex mixing of the trivial probability distributions $f_k(x_i) = \delta_{ik}$, represented by the unit vectors of \mathbb{R}^N . Finally, it's also interesting to consider the inner product induced on $\mathbb{R}^{\mathcal{X}}$ by the Euclidean one of \mathbb{R}^N . Let p be a probability distribution and A a random variable, then

$$p \cdot A = \sum_{x \in \mathcal{X}} p(x)A(x) = \mathbb{E}_p[A] \quad (1.6)$$

1.1.2 Statistical models and manifolds

We call an n-dimensional *statistical model* on \mathcal{X} a family of probability distributions that are globally parametrized by n real-valued variables. Formally this is a subset \mathcal{S} of the space of probability distributions with an invertible function $\psi: \mathcal{S} \rightarrow \Xi \subseteq \mathbb{R}^n$, so that we may write

$$\mathcal{S} = \left\{ p_{\xi} \mid \exists \xi = (\xi^{(1)}, \dots, \xi^{(n)}) \in \Xi: p_{\xi} = \psi^{-1}(\xi) \right\} \quad (1.7)$$

where $p_{\xi}(x)$ may be equivalently written as $p(x; \xi)$ or $p(x; \xi^{(1)}, \dots, \xi^{(n)})$. This definition of a statistical model reflects the act of hypothesizing an underlying model, that may depend on some parameters, for the generation of the random variable's samples. Then only a subset, here represented by \mathcal{S} , of all the possible probability distributions is considered as a candidate of the underlying probability distribution, and every candidate probability distribution is identified uniquely by the corresponding parameters, here represented by ξ .

We now introduce some additional requirements to statistical models so that we may define well-behaved manifolds from them. Firstly we regard S

as a subset of \mathcal{A}_1 equipped with the topology induced by the standard one of \mathbb{R}^N . Then we assume that

$$\begin{aligned}\Xi &\text{ is an open set} \\ \psi &\text{ is a } C^\infty \text{ diffeomorphism from } \mathcal{S} \text{ to } \Xi\end{aligned}\tag{1.8}$$

This allows us to differentiate the probability distributions with respect to the parameters so that $\partial_i p(x; \xi)$ is well defined, where we wrote $\partial_i := \frac{\partial}{\partial \xi^{(i)}}$. These conditions also imply that the pair \mathcal{S} and ψ form a chart of \mathcal{S} . Then for any another statistical model on \mathcal{S} with parametrization $\psi': \mathcal{S} \rightarrow \Xi' \subseteq \mathbb{R}^n$ that follows eq. (1.8), the composed function $\psi' \circ \psi^{-1}: \Xi \rightarrow \Xi'$ will be a C^∞ diffeomorphism. By considering all the possible parametrizations of this kind we may treat \mathcal{S} as a C^∞ -differentiable manifold, where statistical models are the charts and the different parametrizations are the coordinate systems; we call manifolds like these *statistical manifolds*.

From our definitions, it is clear that the maximal dimension of a model is $n = N - 1$ and that every statistical manifold is a submanifold of

$$\mathcal{P} := \{p \in \mathbb{R}^{\mathcal{X}} \mid p(x) > 0 \quad \forall x \in \mathcal{X} \quad \text{and} \quad \sum_{x \in \mathcal{X}} p(x) = 1\}\tag{1.9}$$

that we call the *manifold of probability distributions*. Notice that \mathcal{P} is the interior of the space of probability distributions, this is because from our definitions follows that every $(N - 1)$ -dimensional statistical manifold must be an open subset of \mathcal{A}_1 . ([Examples?](#))

1.1.3 The tangent space and its representations

We will now study tangent vectors of statistical manifolds looking for useful statistical interpretations of them. To do this we will use the fact that, as explained in section 1.1.1, \mathcal{P} can be embedded in the space of random variables $\mathbb{R}^{\mathcal{X}}$. Then we can try to also embed the tangent spaces in $\mathbb{R}^{\mathcal{X}}$ in some meaningful ways, thus linking tangent vectors and random variables.

The mixture representation

Since \mathcal{P} is an open subset of the affine space \mathcal{A}_1 we can naturally identify the tangent space at every point with the displacement vector space $\mathcal{A}_0 := \{A \in \mathbb{R}^{\mathcal{X}} \mid \sum_{x \in \mathcal{X}} A(x) = 0\}$. This is the natural embedding of $T_p(\mathcal{P})$ that arises from the trivial embedding of \mathcal{P} in $\mathbb{R}^{\mathcal{X}}$, in fact for any $X \in T_p(\mathcal{P})$ we can define

$$X^{(m)}(x) := X(p(x))\tag{1.10}$$

then by considering a parametrization $\{\xi^{(i)}\}$ of \mathcal{P} and its relative coordinate basis $\{\partial_i\}$ we have that

$$\partial_i^{(m)}(x) = \partial_i p(x; \xi) \in \mathcal{A}_0 \quad (1.11)$$

since

$$\sum_{x \in \mathcal{X}} \partial_i p(x; \xi) = \partial_i \sum_{x \in \mathcal{X}} p(x; \xi) = 0 \quad (1.12)$$

Finally, from eq. (1.8) follows that $\partial_i p(x; \xi)$ are $N - 1$ linearly independent functions and thus

$$X^{(m)} \leftrightarrow X \quad (1.13)$$

is an isomorphism and

$$T_p(\mathcal{P}) \sim T_p^{(m)}(\mathcal{P}) := \{X^{(m)} \mid X \in T_p(\mathcal{P})\} = \mathcal{A}_0 \quad \forall p \in \mathcal{P} \quad (1.14)$$

We call $X^{(m)}$ the *mixture representation* or *m-representation* of X .

The exponential representation

Since $T_p(\mathcal{P})$ is an $(N - 1)$ -dimensional vector space, for every p we may try to identify it to the subspace of $\mathbb{R}^{\mathcal{X}}$ orthogonal to p with respect to the inner product defined in eq. (1.6). This is interesting given the statistical meaning of the inner product between a generic element of $\mathbb{R}^{\mathcal{X}}$ and a probability distribution. For every $p \in \mathcal{P}$ the orthogonal space is

$$\mathcal{A}_p^\perp := \{A \in \mathbb{R}^{\mathcal{X}} \mid p \cdot A = E_p[A] = 0\} \quad (1.15)$$

that is the space of random variables with null expectation value when the underlying probability distribution is p .

Now we wish to find a natural isomorphism between $T_p(\mathcal{P})$ and \mathcal{A}_p^\perp . One way to do this that will prove to be useful is to consider the following alternative embedding of \mathcal{P} in $\mathbb{R}^{\mathcal{X}}$

$$p \mapsto \ln p \in \mathbb{R}^{\mathcal{X}} \quad (1.16)$$

then, for any $X \in T_p(\mathcal{P})$ we can define

$$X^{(e)}(x) := X(\ln p(x)) = \frac{X(p(x))}{p(x)} \quad (1.17)$$

and by considering a parametrization $\{\xi^{(i)}\}$ of \mathcal{P} and its relative coordinate basis $\{\partial_i\}$ we have that

$$\partial_i^{(e)}(x) = \partial_i \ln p(x; \xi) \in \mathcal{A}_p^\perp \quad (1.18)$$

since

$$\mathbb{E}_p[\partial_i \ln p(x; \xi)] = \sum_{x \in \mathcal{X}} p(x; \xi) \frac{\partial_i p(x; \xi)}{p(x; \xi)} = \sum_{x \in \mathcal{X}} \partial_i p(x; \xi) = 0 \quad (1.19)$$

It's easy to prove that the linear independence of $\partial_i \ln p(x; \xi)$ follows from the one of $\partial_i p(x; \xi)$ and thus

$$X^{(e)} \leftrightarrow X \quad (1.20)$$

is an isomorphism and

$$T_p(\mathcal{P}) \sim T_p^{(e)}(\mathcal{P}) := \{X^{(e)} \mid X \in T_p(\mathcal{P})\} = \mathcal{A}_p^\perp \quad \forall p \in \mathcal{P} \quad (1.21)$$

We call $X^{(m)}$ the *exponential representation* or *e-representation* of X .

From their definitions, we have that the two representations of a tangent vector $X \in T_p(\mathcal{P})$ are related as follows

$$X^{(m)}(x) = X^{(e)}(x)p(x) \quad (1.22)$$

and that while $T_p^{(m)}(\mathcal{P})$ is the same for every p , $T_p^{(e)}(\mathcal{P})$ varies since the set of random variables with null expectation value will differ depending on the underlying probability distribution.

1.2 The information metric

1.2.1 Relative entropy

Given a statistical manifold, we may ask ourselves if a certain metric can be naturally defined on it. Such a metric would give rise to a Riemannian connection and consequently to a geodesic distance between elements of the manifold. For this reason, we should first find a statistical meaning to the notion of distance between probability distributions and only then try to find a metric coherent with it.

One natural way to proceed is to consider how hard it is to distinguish a probability distribution from another one by extracting some samples. More precisely let's assume that a random process has an underlying probability distribution q and that N_s samples are generated. Then we can consider the probability that the resulting frequencies f_i of the samples correspond to the probabilities p_i of another probability distribution p .

For simplicity, let's consider the $N = 2$ case, i.e. the case of binomial distributions. Let $\mathbf{q} = (t, 1 - t)$ and $\mathbf{p} = (r, 1 - r)$ be two probability

distributions. Then if N_s samples are drawn with underlying probability q , the probability $P_{N_s}(\mathbf{p})$ that the obtained frequencies correspond to \mathbf{p} is given by

$$P_{N_s}(\mathbf{p}) = \binom{N_s}{N_s \cdot r} t^{N_s \cdot r} (1-t)^{N_s \cdot (1-r)}$$

then by assuming $r \neq 0, 1$ and using Stirling's formula, we obtain the following asymptotic behavior for $N_s \rightarrow \infty$

$$P_{N_s}(\mathbf{p}) \sim \exp \left\{ -N_s \left[r \ln \left(\frac{r}{t} \right) + (1-r) \ln \left(\frac{1-r}{1-t} \right) \right] \right\} \quad (1.23)$$

so the probability decreases exponentially with N_s times the factor in the square parenthesis. This factor only depends on the probability distributions p and q , and one may recognize it from statistics as the *relative entropy* of the two distributions. For a generic finite sample space \mathcal{X} , the relative entropy is defined as

$$S(p \parallel q) := \sum_{x \in \mathcal{X}} p(x) \ln \left(\frac{p(x)}{q(x)} \right) \quad (1.24)$$

More generally it can be shown that the following theorem holds.

Theorem 1 (Sanov's Theorem). *Let $\mathcal{E} \subset \mathcal{P}$ be a closed set of probability distributions without isolated points. Then if N_s samples are drawn with underlying probability distribution $q \in \mathcal{P}$, the probability $P_{N_s}(\mathcal{E})$ that the obtained frequencies correspond to an element in \mathcal{E} has the following asymptotic behavior*

$$P_{N_s}(\mathcal{E}) \sim e^{-N_s S(p_* \parallel q)} \quad \text{for } N_s \rightarrow \infty \quad (1.25)$$

where p_* is the element of \mathcal{E} for which $S(p_* \parallel q)$ is smallest.

Roughly speaking, this shows that the greater the relative entropy $S(p \parallel q)$ the faster the probability of obtaining frequencies in a small neighborhood of p decreases with the number of samples drawn with underlying probability q . In this view, relative entropy can serve as a kind of distance between probability distributions, but with some caveats. From the definition in eq. (1.24) relative entropy has the following properties

$$S(p \parallel q) \geq 0 \quad \forall p, q \in \mathcal{P} \quad (1.26)$$

$$S(p \parallel q) = 0 \iff p = q \quad (1.27)$$

but it is not symmetric and it doesn't follow the triangle inequality, so it is not a metric distance. We may ask ourselves if the asymmetry is an accident of our definition of relative entropy or if it is inherent in the distinguishability of probability distributions. The latter turns out to be true, as shown by the following example.

Example. Consider two coins, one fair and one with heads on both sides. We want to pick one and guess which one it is just by tossing it multiple times. Clearly, the game is not symmetric in the choice of the coin; in fact, if we pick the fair coin the first time we will get a tail we will be sure that we picked the fair one, while if we pick the double-head one we will only get heads but this result will always be also compatible with a fair coin that, by chance, is only giving heads.

This game is precisely a problem of distinguishability of probability distributions. In fact, we have the $N = 2$ sample space and two probability distributions: $\mathbf{p} = (0.5, 0.5)$ (the fair coin) and $\mathbf{q} = (1, 0)$ (the double-head coin), so the two relative entropies are $S(p \parallel q) \rightarrow \infty$ and $S(q \parallel p) = \ln 2$. If we pick the fair coin for large N_s the obtained frequencies will approach \mathbf{p} ; then we consider the probability of obtaining these frequencies if the underlying distribution was q . This probability is identically 0, and it is coherent with Sanov's theorem since $S(p \parallel q) \rightarrow \infty$ and so $P_{N_s}(p) \sim e^{-N_s \cdot \infty} = 0$. Otherwise, if we pick the double-head coin the frequencies will always match exactly \mathbf{q} , and the probability of getting this if the underlying distribution was q is 0.5^{N_s} . This is coherent with Sanov's theorem since $S(q \parallel p) = \ln 2$ and so $P_{N_s}(q) \sim e^{-N_s \cdot \ln 2} = 0.5^{N_s}$.

Even though relative entropy doesn't follow triangle inequality, it does follow a generalization of the Pythagorean theorem stated as follows

Theorem 2 (Generalized Pythagorean theorem). *Let $\mathcal{E} \subset \mathcal{P}$ be a convex set and consider $p \in \mathcal{E}$ and $q \in \mathcal{P} \setminus \mathcal{E}$. Then*

$$S(p \parallel q) \geq S(p \parallel p_*) + S(p_* \parallel q) \quad (1.28)$$

where p_* is the element of $\partial\mathcal{E}$ for which $S(p_* \parallel q)$ is smallest.

This is a generalization of the Pythagorean theorem in the sense that if it was stated in terms of the Euclidean distance squared, the angle between $\overline{pp_*}$ and $\overline{p_*q}$ would be obtuse and so eq. (1.28) would be the corollary of the Pythagorean theorem for obtuse triangles. This suggests that the relative entropy may be regarded as an asymmetric distance squared and as we will see this is enough to define a metric on the manifold.

1.2.2 Squared Riemannian distances

Now that we have some notion of distance, we shall explore how to define a coherent metric on the manifold. First, we shall study this for a Riemannian distance.

Let (M, g) be a Riemannian manifold where g is the metric tensor, then consider a point $p \in M$. We define the *exponential map* in p as follows

$$\text{Exp}_p : T_p M \rightarrow M, \quad \text{Exp}_p(\mathbf{v}) := \gamma_{\mathbf{v}}(1) \quad (1.29)$$

where $\mathbf{v} \in T_p M$ and $\gamma_{\mathbf{v}} : [0, 1] \rightarrow M$ is the unique geodesic tangent to v in p , i.e. satisfying $\gamma_{\mathbf{v}}(0) = p$ and $\gamma'_{\mathbf{v}}(0) = \mathbf{v}$. In general, this map will be well-defined only from a neighborhood of the origin of $T_p M$ to a neighborhood of p , since only locally the uniqueness of the geodesic curve is guaranteed. By eventually further restricting the neighborhood, this map will also be 1-1 since locally the geodesic curves don't cross.

Then in this neighborhood, we have the inverse of the exponential map that maps $q \mapsto \mathbf{v}_q \in T_p M$ so that $\gamma_{\mathbf{v}_q}(1) = q$. Since $\gamma_{\mathbf{v}_q}$ is the only geodesic connecting p and q , the geodesic distance between them will be

$$L(p, q) = \int_0^1 \sqrt{g_{\gamma_{\mathbf{v}_q}(\lambda)}(\gamma'_{\mathbf{v}_q}(\lambda), \gamma'_{\mathbf{v}_q}(\lambda))} d\lambda \quad (1.30)$$

Because $\gamma_{\mathbf{v}_q}$ is a geodesic, by definition we have that $\gamma'_{\mathbf{v}_q}(\lambda)$ is parallel transported along the curve and so

$$g_{\gamma_{\mathbf{v}_q}(\lambda)}(\gamma'_{\mathbf{v}_q}(\lambda), \gamma'_{\mathbf{v}_q}(\lambda)) = g_p(\mathbf{v}_q, \mathbf{v}_q) \quad \forall \lambda \in [0, 1] \quad (1.31)$$

Finally then, we get

$$L(p, q) = \int_0^1 \sqrt{g_p(\mathbf{v}_q, \mathbf{v}_q)} d\lambda = \sqrt{g_p(\mathbf{v}_q, \mathbf{v}_q)} = \|\mathbf{v}_q\| \quad (1.32)$$

and so $\hat{\mathbf{v}}_q = \mathbf{v}_q \setminus L(p, q)$.

Now chose a vector $\mathbf{dp} \in T_p M$ and let $q = \text{Exp}_p(\mathbf{dp})$. Then, let $\Gamma_{\hat{\mathbf{dp}}}$ be the unique geodesic curve such that $\Gamma_{\hat{\mathbf{dp}}}(0) = p$ and $\Gamma'_{\hat{\mathbf{dp}}}(0) = \hat{\mathbf{dp}}$. Clearly, this is the following reparametrization of $\gamma_{\mathbf{dp}}$

$$\Gamma_{\hat{\mathbf{dp}}}(l) = \gamma_{\mathbf{dp}}\left(\frac{l}{\|\mathbf{dp}\|}\right) = \gamma_{\mathbf{dp}}\left(\frac{l}{L(p, q)}\right) \quad (1.33)$$

and so $q = \Gamma_{\hat{\mathbf{dp}}}(L(p, q))$.

Now let $\{x^{(i)}\}$ be a coordinate system for the neighborhood; from the Taylor expansion of $\Gamma_{\hat{\mathbf{dp}}}^{(i)}(l)$ in $l = 0$ we get

$$\Gamma_{\hat{\mathbf{dp}}}^{(i)}(l) = \Gamma_{\hat{\mathbf{dp}}}^{(i)}(0) + \left. \frac{d\Gamma_{\hat{\mathbf{dp}}}^{(i)}(l)}{dl} \right|_{l=0} \cdot l + O(l^2) \quad (1.34)$$

then, from $\Gamma_{\hat{\mathbf{d}}\mathbf{p}}(0) = p$, $\Gamma'_{\hat{\mathbf{d}}\mathbf{p}}(0) = \hat{\mathbf{d}}\mathbf{p}$ and $q = \Gamma_{\hat{\mathbf{d}}\mathbf{p}}(L(p, q))$ we get that

$$q^{(i)} = p^{(i)} + \hat{dp}^i L(p, q) + O(L^2(p, q)) = p^{(i)} + dp^i + O(\|\mathbf{d}\mathbf{p}\|^2) \quad (1.35)$$

for every q in the image of $\Gamma_{\hat{\mathbf{d}}\mathbf{p}}$. Then, combining eq. (1.32) and eq. (1.35) we get

$$L^2(\{p^{(i)}\}, \{p^{(i)} + dp^i\}) \rightarrow L^2(p, q) = g_{ij} dp^i dp^j \quad \text{for } dp^i \rightarrow 0 \quad (1.36)$$

We define $L_p^2 : M \rightarrow [0, +\infty)$, $L_p^2(q) := L^2(p, q)$ and consider its Taylor expansion in $q = p$

$$L_p^2(\{p^{(i)} + dp^i\}) = \frac{1}{2} \left[\partial_i \partial_j L_p^2(q) \right]_{q=p} dp^i dp^j + O(dp^3) \quad (1.37)$$

where the first derivative vanishes because p is a minimum of L_p^2 . Then we get

$$g_{ij}^{(p)} = \frac{1}{2} \left[\partial_i \partial_j L_p^2(q) \right]_{q=p} \quad (1.38)$$

so the metric tensor in p is proportional to the Hessian matrix of L_p^2 in p . This is possible because the Hessian matrix of any C^2 function of a differentiable manifold evaluated in a critical point is a $\binom{0}{2}$ tensor.

We have thus found a way to recover the metric tensor from its squared Riemannian distance.

1.2.3 Divergences

As argued in the last paragraph of section 1.2.1, relative entropy has some properties of squared distances and so we may try to find a metric tensor coherent with it as we did in section 1.2.2.

Let M be a differentiable manifold and $D(\cdot \| \cdot) : M \times M \rightarrow [0, +\infty)$ a C^2 function (possibly asymmetric) satisfying

$$D(p \| q) \geq 0 \quad \text{and} \quad D(p \| q) = 0 \iff p = q \quad \forall p, q \in M \quad (1.39)$$

Then, given a coordinate system $\{\xi^{(i)}\}$ on M we have that every pair of points $(q, q') \in M \times M$ has coordinates $(\{\xi^{(i)}\}, \{\xi^{(i)'}\})$ and we use the following notation for partial derivatives in one of the two terms on the diagonal (p, p)

$$\begin{aligned} D[\partial_i \| \cdot] &: p \mapsto [\partial_i D(q \| q')]_{(q,q')=(p,p)} \\ D[\cdot \| \partial_i] &: p \mapsto [\partial'_i D(q \| q')]_{(q,q')=(p,p)} \end{aligned}$$

From the fact that the diagonal (p, p) is a constant surface of minima of D follows that

$$D[\partial_i \parallel \cdot] = D[\cdot \parallel \partial_i] \equiv 0 \quad (1.40)$$

so the diagonal is also a constant surface of the derivatives of D . Then by further deriving parallel to the diagonal, we get

$$\begin{aligned} (\partial_i + \partial'_i)D[\cdot \parallel \partial_j] &= D[\cdot \parallel \partial_i \partial_j] + D[\partial_i \parallel \partial_j] \equiv 0 \\ (\partial_j + \partial'_j)D[\partial_i \parallel \cdot] &= D[\partial_i \partial_j \parallel \cdot] + D[\partial_i \parallel \partial_j] \equiv 0 \end{aligned}$$

where we used the fact that since D is C^2 we can swap second-order derivatives. Finally, we get

$$D[\partial_i \partial_j \parallel \cdot] = D[\cdot \parallel \partial_i \partial_j] = -D[\partial_i \parallel \partial_j] =: g_{ij}^{(D)} \quad (1.41)$$

where from the fact that the diagonal is a surface of minima, it follows that the previous expression defines a (symmetric) positive semi-definite tensor. From eq. (1.40) and eq. (1.41) it follows that denoting

$$D_p^{(R)} : q \mapsto D(p \parallel q) \quad \text{and} \quad D_p^{(L)} : q \mapsto D(q \parallel p)$$

to second order, we get

$$D_p^{(R)}(q) = \frac{1}{2}g_{ij}^{(D)}d\xi^i d\xi^j + O(d\xi^2) \quad \text{and} \quad D_p^{(L)}(q) = \frac{1}{2}g_{ij}^{(D)}d\xi^i d\xi^j + O(d\xi^2)$$

where $d\xi^i := \xi_p^{(i)} - \xi_q^{(i)}$

and so to the lowest order, the asymmetry is not present.

Finally, if $g_{ij}^{(D)}$ is positive definite we say that D is a *divergence*, then $\frac{1}{2}g_{ij}^{(D)}$ defines a metric tensor and so a unique Riemannian structure on the manifold. The induced squared Riemannian distance coincides at the lowest order near the diagonal with the divergence.

1.2.4 The Fisher information metric

We now go back to the relative entropy, we report the definition given in section 1.2.1

$$S(p \parallel q) := \sum_{x \in \mathcal{X}} p(x) \ln \left(\frac{p(x)}{q(x)} \right)$$

this is a C^2 function of $\mathcal{P} \times \mathcal{P}$ and it follows eq. (1.39). So for any model \mathcal{S} with parameters $\{\xi^{(i)}\}$ we can define

$$g_{ij}^{(S)} = S[\cdot \parallel \partial_i \partial_j] = \partial'_i \partial'_j \sum_{x \in \mathcal{X}} p(x; \xi) \ln \left(\frac{p(x; \xi)}{q(x; \xi')} \right) \Big|_{\xi=\xi'} \quad (1.42)$$

And so we have

$$g_{ij}^{(S)} = - \sum_{x \in \mathcal{X}} p(x) \partial_i \partial_j \ln p(x; \xi) = -\mathbb{E}_p[\partial_i \partial_j \ln p(x; \xi)] \quad (1.43)$$

and equivalently

$$g_{ij}^{(S)} = \mathbb{E}_p \left[\frac{1}{p^2(x)} \partial_i p(x; \xi) \partial_j p(x; \xi) \right] = \mathbb{E}_p[\partial_i \ln p(x; \xi) \partial_j \ln p(x; \xi)] \quad (1.44)$$

$$= \sum_{x \in \mathcal{X}} \frac{\partial_i p(x; \xi) \partial_j p(x; \xi)}{p(x)} \quad (1.45)$$

where we used the fact that

$$\sum_{x \in \mathcal{X}} \partial_i \partial_j p(x; \xi) = \partial_i \partial_j \sum_{x \in \mathcal{X}} p(x; \xi) = 0 \quad (1.46)$$

Since we know that any $g_{ij}^{(D)}$ is positive semi-definite, $g_{ij}^{(S)}$ will be positive definite if and only if it is invertible. It can be easily shown that if the functions $\partial_i p(x; \xi)$ are linearly independent, then $g_{ij}^{(S)}$ is invertible and thus positive definite (**proof?**). Thus, relative entropy is a divergence and in fact, it is also known as *Kullback-Leibler divergence* or *information divergence*.

Finally, then, $G_F := \{g_{ij}^{(S)}\}$ defines a metric tensor at every point and it is known as the *Fisher information metric*. This is, up to a constant factor, the unique metric induced by the relative entropy, and it plays a focal role in the geometrical modeling and interpretation of statistics.

1.2.5 The geometry of \mathcal{P}

The Fisher metric defines the inner product $\langle \cdot, \cdot \rangle_p$ between tangent vectors of a point $p \in \mathcal{P}$. Let us now express this inner product through the representations we defined in section 1.1.3. Given two tangent vectors $X, Y \in T_p \mathcal{P}$ from eq. (1.44) we find that

$$\langle X, Y \rangle_p = \mathbb{E}_p[X^{(e)} Y^{(e)}] \quad (1.47)$$

while from eq. (1.45) we get

$$\langle X, Y \rangle_p = \sum_{x \in \mathcal{X}} \frac{X^{(m)}(x) Y^{(m)}(x)}{p(x)} \quad (1.48)$$

$$= \sum_{x \in \mathcal{X}} X^{(m)}(x) Y^{(e)}(x) = \sum_{x \in \mathcal{X}} X^{(e)}(x) Y^{(m)}(x) \quad (1.49)$$

These expressions will prove to be very useful in section 1.3.

We notice that in neither representation the inner product is the Euclidean one induced by $\mathbb{R}^{\mathcal{X}}$ on the respective embeddings. For such a representation we would have that

$$\langle X, Y \rangle_p = \sum_{x \in \mathcal{X}} X^{(0)}(x) Y^{(0)}(x) \quad (1.50)$$

then it's easy to guess that the embedding

$$p \mapsto 2\sqrt{p} =: p_{(0)} \quad (1.51)$$

is the one whose representation of the tangent spaces

$$X^{(0)} := X(2\sqrt{p}) = \frac{X(p)}{\sqrt{p}}, \quad X \in T_p(\mathcal{P}) \quad (1.52)$$

follows eq. (1.50). This means that the information geometry of \mathcal{P} , i.e. the geometry induced on it by the Fisher metric, is that of an N -dimensional round sphere (of radius 2) since

$$\sum_{x \in \mathcal{X}} p(x) = 1 \implies \sum_{x \in \mathcal{X}} p_{(0)}^2(x) = 4 \quad (1.53)$$

(Talk about e and m connections?)

1.3 Parameter estimation

1.3.1 Unbiased estimators

Consider a random process and an n -dim statistical model $\mathcal{S} = \{p_\xi \mid \xi \in \Xi\}$ of it, as defined in section 1.1.2; it is often the case that from a measured sample $x \in \mathcal{X}$ we want to estimate the parameters ξ of the underlying probability distribution, that we assume to be in \mathcal{S} .

The estimation is represented by a function

$$\hat{\xi} = (\hat{\xi}^{(1)}, \dots, \hat{\xi}^{(n)}) : \mathcal{X} \rightarrow \Xi \subseteq \mathbb{R}^n \quad (1.54)$$

that we call *estimator*. Each component $\hat{\xi}^{(i)}$ is a random variable, and we say that $\hat{\xi}$ is an *unbiased estimator* if

$$E_{p_\xi} [\hat{\xi}] = \left(E_{p_\xi} [\hat{\xi}^{(1)}], \dots, E_{p_\xi} [\hat{\xi}^{(n)}] \right) = \xi \quad \forall \xi \in \Xi \quad (1.55)$$

i.e. if for each $p_\xi \in \mathcal{S}$ the expectation value of the estimator is the correct parameter ξ .

Then for an unbiased estimator, we may represent the deviation from the true parameters with the variance-covariance matrix of the estimator $V_\xi [\hat{\xi}] := \{v_\xi^{ij}\}$ where

$$v_\xi^{ij} := \text{Cov}_{p_\xi} [\hat{\xi}^{(i)}, \hat{\xi}^{(j)}] = \mathbb{E}_{p_\xi} \left[(\hat{\xi}^{(i)}(x) - \xi^i)(\hat{\xi}^{(j)}(x) - \xi^j) \right] \quad (1.56)$$

In particular, the elements on the diagonal are the variances of the components of the estimator. (Talk about the covariance ellipse?)

1.3.2 Variance and expectation value

For a generic random variable $A \in \mathbb{R}^\mathcal{X}$ we may define a real function $\mathbb{E}[A]$ on \mathcal{P} that maps every probability distribution p to the expectation value of A when the sample is generated with p as underlying probability distribution

$$\mathbb{E}[A]: \mathcal{P} \rightarrow \mathbb{R} \quad p \mapsto \mathbb{E}_p[A] \quad (1.57)$$

Since this is a function of the manifold \mathcal{P} , at every point p we may consider its differential $(d\mathbb{E}[A])_p$. This is the element of the cotangent space $T_p^*(\mathcal{P})$ such that for any tangent vector $X \in T_p(\mathcal{P})$ we have

$$(d\mathbb{E}[A])_p(X) = X(\mathbb{E}[A]) \quad (1.58)$$

Also, because a metric is defined on \mathcal{P} we have a natural isomorphism between tangent and cotangent vectors, thus the gradient of $\mathbb{E}[A]$ is the tangent vector defined by

$$\langle (\text{grad}\mathbb{E}[A])_p, X \rangle_p = (d\mathbb{E}[A])_p(X) = X(\mathbb{E}[A]) \quad \forall X \in T_p(\mathcal{P}) \quad (1.59)$$

We now state and prove the following theorem that relates the deviation of a random variable to the gradient of its expectation value

Theorem 3. *For any random variable $A \in \mathbb{R}^\mathcal{X}$ we have that*

$$(\text{grad}\mathbb{E}[A])_p^{(e)} = A - \mathbb{E}_p[A] \quad \forall p \in \mathcal{P} \quad (1.60)$$

where the gradient is the dual tangent vector of the differential with respect to the Fisher metric.

Proof. For every $X \in T_p$ we have

$$\begin{aligned} X(\mathbb{E}[A]) &= \sum_{x \in \mathcal{X}} X(p(x))A(x) = \sum_{x \in \mathcal{X}} X^{(m)}(x)A(x) \\ &= \mathbb{E}_p[X^{(e)}A] = \mathbb{E}_p[X^{(e)}(A - \mathbb{E}_p[A])] \end{aligned}$$

where in the last equation we used the fact that $\mathbb{E}_p[X^{(e)}] = 0$. We notice that

$$\mathbb{E}_p[A - \mathbb{E}_p[A]] = 0 \implies (A - \mathbb{E}_p[A]) \in T_p^{(e)}(\mathcal{P})$$

and so there must exist a tangent vector $Y \in T_p(\mathcal{P})$ such that $Y^{(e)} = (A - \mathbb{E}_p[A])$. Then

$$X(\mathbb{E}[A]) = \mathbb{E}_p[X^{(e)}Y^{(e)}] = \langle X, Y \rangle_p$$

and so from eq. (1.59) we have that $Y = (\text{grad}\mathbb{E}[A])_p$ and so

$$(\text{grad}\mathbb{E}[A])_p^{(e)} = A - \mathbb{E}_p[A]$$

□

We also get the following

Corollary 3.1. *For any random variable A*

$$\mathbb{V}_p[A] = \|(\text{d}\mathbb{E}[A])_p\|_p^2 \quad (1.61)$$

Proof. Follows from theorem 3 noticing that

$$\begin{aligned} \|(\text{d}\mathbb{E}[A])_p\|_p^2 &= \langle (\text{grad}\mathbb{E}[A])_p, (\text{grad}\mathbb{E}[A])_p \rangle_p \\ &= \mathbb{E}_p[(A - \mathbb{E}_p[A])^2] \end{aligned}$$

□

Let now \mathcal{S} be an n -dim statistical manifold, since it is a submanifold of \mathcal{P} we have that $T_p(\mathcal{S})$ is a linear subspace of $T_p(\mathcal{P})$ for every $p \in \mathcal{S}$. Then the gradient of the restriction on \mathcal{S} of a function of \mathcal{P} is its orthogonal projection on $T_p(\mathcal{S})$. In particular, we have that

$$T_p(\mathcal{P}) = T_p(\mathcal{S}) \oplus T_p(\mathcal{S})^\perp \quad (1.62)$$

and so we may uniquely decompose

$$(\text{grad}\mathbb{E}[A])_p = v_{\parallel} + v_{\perp} \quad v_{\parallel} \in T_p(\mathcal{S}), v_{\perp} \in T_p(\mathcal{S})^\perp \quad (1.63)$$

then we find that $\forall X \in T_p(\mathcal{S})$

$$\langle (\text{gradE}[A]|_{\mathcal{S}})_p, X \rangle_p = X(\text{E}[A]) = \langle v_{\parallel} + v_{\perp}, X \rangle_p = \langle v_{\parallel}, X \rangle_p \quad (1.64)$$

and so $(\text{gradE}[A]|_{\mathcal{S}})_p = v_{\parallel}$.

Finally, we have the following theorem relating the variance of a random variable and the sensitivity of its expectation value to the changes in the model parameters

Theorem 4. *Given a statistical manifold S , for any random variable A we have that*

$$V_p[A] \geq \|(\text{dE}[A]|_{\mathcal{S}})_p\|_p^2 \quad (1.65)$$

where the equality holds if and only if

$$A - \text{E}_p[A] \in T_p^{(e)}(\mathcal{S}) \quad (1.66)$$

Proof. Follows immediately from Corollary 3.1 and eq. (1.64) □

1.3.3 The Cramér-Rao bound

We are now in the position to state and prove an important result of parameter estimation theory

Theorem 5 (Cramér-Rao bound). *Let $\mathcal{S} = \{p_{\xi} \mid \xi \in \Xi\}$ be an n -dim statistical model of \mathcal{P} . Then, for any unbiased estimator $\hat{\xi}$ the variance-covariance matrix $V_{\xi}[\hat{\xi}]$ satisfies*

$$V_{\xi}[\hat{\xi}] \geq G_F^{-1}(p_{\xi}) \quad (1.67)$$

in the sense that $V_{\xi}[\hat{\xi}] - G_F^{-1}(p_{\xi})$ is positive semi-definite.

Proof. Let $A = c_i \hat{\xi}^{(i)}$ where c is an arbitrary element of \mathbb{R}^n . Then A is a random variable with $\text{E}_{p_{\xi}}[A] = c_i \xi^i$ and

$$\begin{aligned} V_{p_{\xi}}[A] &= \text{E}_{p_{\xi}}[(c_i \hat{\xi}^{(i)}(x) - c_i \xi^i)(c_j \hat{\xi}^{(j)}(x) - c_j \xi^j)] \\ &= \text{E}_{p_{\xi}}[c_i(\hat{\xi}^{(i)}(x) - \xi^i)(\hat{\xi}^{(j)}(x) - \xi^j)c_j] \\ &= c_i v_{\xi}^{ij} c_j \end{aligned}$$

Then letting $p = p_{\xi}$ from theorem 4 we get

$$c_i v_{\xi}^{ij} c_j \geq \|(\text{dE}[A]|_{\mathcal{S}})_p\|_p^2$$

where, in the coordinate basis of $\{\xi^{(i)}\}$

$$\begin{aligned}\|(dE[A]|_S)_p\|_p^2 &= (\partial_i E[A])_p g^{ij}(p) (\partial_j E[A])_p \\ &= c_i g^{ij}(p) c_j\end{aligned}$$

and so, finally

$$c_i(v_\xi^{ij} - g^{ij}(p_\xi))c_j \geq 0$$

□

An unbiased estimator that saturates eq. (1.67) is called an *efficient estimator* and is the best unbiased estimator in the sense that its variance is minimum between all unbiased estimators. It's important to notice that an efficient estimator doesn't always exist. (**talk about asymptotically efficient estimators?**)

This result shows that the efficiency with which we can infer the underlying probability distribution of a process is deeply linked with the information geometry of the model. (**Examples**)

Chapter 2

Geometry of quantum states

2.1 Quantum mechanics

2.1.1 The postulates of quantum mechanics

Quantum mechanics is a fundamental theory that provides a mathematical framework for modeling the states and evolution of physical systems, as well as predicting the results of observations. In this section, we will state the postulates of quantum mechanics, roughly following the treatment in ??.

Postulate 0 (The Hilbert space). *Any isolated physical system is associated with a complex Hilbert space i.e. a, possibly infinite-dimensional, complex vector space \mathcal{H} with an inner product $\langle \cdot, \cdot \rangle$ that is also complete with respect to the metric induced by the inner product.*

If the Hilbert space is finite-dimensional with dimension N , then it is isomorphic to \mathbb{C}^N with a hermitian form as the inner product. In what follows we will only study systems that are associated with a finite-dimensional Hilbert space.

The chosen inner product allows us to define a canonical isomorphism between \mathcal{H} and its dual vector space \mathcal{H}^* , where for every $\mathbf{Z} \in \mathcal{H}$ its dual functional is defined as

$$\mathbf{Z}^* \equiv f_{\mathbf{Z}} : \mathcal{H} \rightarrow \mathbb{C} \quad f_{\mathbf{Z}}(\mathbf{X}) = \langle \mathbf{Z}, \mathbf{X} \rangle \quad \forall \mathbf{X} \in \mathcal{H} \quad (2.1)$$

We also define tensors T of type $\binom{p}{q}$ as multilinear maps

$$T : \mathcal{H}^{*p} \times \mathcal{H}^q \rightarrow \mathbb{C} \quad (2.2)$$

where vectors can be identified with tensors of type $\binom{1}{0}$ and dual vectors with tensors of type $\binom{0}{1}$. Operators, which are linear maps from \mathcal{H} to itself, can

be identified with $\binom{1}{1}$ tensors. For any operator A , the corresponding tensor T_A is defined as

$$T_A(\mathbf{X}^*, \mathbf{Y}) = \mathbf{X}^*(A\mathbf{Y}) = \langle \mathbf{X}, A\mathbf{Y} \rangle \quad (2.3)$$

The tensor product \otimes between two tensors T and T' of type $\binom{p}{q}$ and $\binom{p'}{q'}$ is a tensor of type $\binom{p+p'}{q+q'}$ defined as

$$\begin{aligned} (T \otimes T')(\mathbf{X}_1^*, \dots, \mathbf{X}_{p+p'}^*, \mathbf{Y}_1, \dots, \mathbf{Y}_{q+q'}) \\ = T(\mathbf{X}_1^*, \dots, \mathbf{X}_p^*, \mathbf{Y}_1, \dots, \mathbf{Y}_q) \cdot T'(\mathbf{X}_{p+1}^*, \dots, \mathbf{X}_{p+p'}^*, \mathbf{Y}_{q+1}, \dots, \mathbf{Y}_{q+q'}) \end{aligned}$$

In particular, given a vector \mathbf{Y} and a dual vector \mathbf{X}^* , their tensor product is the operator

$$A(\mathbf{Z}) = \mathbf{X}^*(\mathbf{Z}) \cdot \mathbf{Y} = \langle \mathbf{X}, \mathbf{Z} \rangle \cdot \mathbf{Y} \quad (2.4)$$

The adjoint of any operator A is the operator A^\dagger such that

$$\langle A^\dagger \mathbf{X}, \mathbf{Y} \rangle = \langle \mathbf{X}, A \mathbf{Y} \rangle \quad \forall \mathbf{X}, \mathbf{Y} \in \mathcal{H} \quad (2.5)$$

It can be shown that this is always well-defined for finite-dimensional Hilbert spaces and that $(A^\dagger)^\dagger = A$. There are two families of operators that will be instrumental to the formulation of the remaining postulates: unitary operators and self-adjoint operators. Unitary operators are defined as operators U that preserve the inner product, i.e.

$$\langle U\mathbf{X}, U\mathbf{Y} \rangle = \langle \mathbf{X}, \mathbf{Y} \rangle \quad \forall \mathbf{X}, \mathbf{Y} \in \mathcal{H} \quad (2.6)$$

From eq. (2.5), it's easy to show that

$$U \text{ is unitary} \iff UU^\dagger = \mathbb{I} \text{ i.e } U^{-1} = U^\dagger \quad (2.7)$$

from which also follows $UU^\dagger = U^\dagger U$. Self-adjoint operators are defined as operators A that are equal to their adjoint, i.e.

$$\langle A \mathbf{X}, \mathbf{Y} \rangle = \langle \mathbf{X}, A \mathbf{Y} \rangle \quad \forall \mathbf{X}, \mathbf{Y} \in \mathcal{H} \quad (2.8)$$

or, equivalently, $A = A^\dagger$. From their definition follows immediately that

$$\langle A \mathbf{X}, \mathbf{X} \rangle = \overline{\langle \mathbf{X}, A \mathbf{X} \rangle} = \overline{\langle A \mathbf{X}, \mathbf{X} \rangle} \in \mathbb{R} \quad \forall \mathbf{X} \in \mathcal{H} \quad (2.9)$$

We can now state the remaining postulates of quantum mechanics for the finite-dimensional case.

Postulate 1 (The state vectors). *Every non-zero vector of the Hilbert space completely characterizes a possible state of the system, we call such vectors state vectors. We thus have that the elements of*

$$\mathcal{H}_0 := \mathcal{H} \setminus \{0\} \quad (2.10)$$

describe all the possible states of the system.

For state vectors, we will also use the Dirac notation writing vectors as $|\psi\rangle$ and their dual as $\langle\psi|$. An exhaustive treatment of this notation may be found in ??

Postulate 2 (Unitary evolution). *The state vectors of a closed system evolve only through unitary transformations of the Hilbert space. That is, the time evolution of any state vector $|\psi(t)\rangle$ is given by*

$$|\psi(t_2)\rangle = U(t_1, t_2) |\psi(t_1)\rangle \quad (2.11)$$

where $U(t_1, t_2)$ is a unitary operator that only depends on t_1 and t_2 .

We may interpret this as requiring that the evolution of a closed system preserves the structure we defined on the set \mathcal{H} , that is the vector space structure and the inner product space structure. Thus we expect the transformations to be invertible, and linear and to preserve the inner product; in this sense, unitary operators are the automorphisms of the Hilbert space.

Postulate 3 (Quantum measurements). *Quantum measurements are described by a collection of pairs $\{(M_x, x)\}$ of measurement operators M_x and outcomes x such that the following completeness equation is satisfied*

$$\sum_x M_x^\dagger M_x = \mathbb{I} \quad (2.12)$$

where \mathbb{I} is the identity operator. Then, given a system in a state described by a state vector $|\psi\rangle$, the probability distribution of the outcomes is

$$p(x) = \frac{\langle\psi|M_x^\dagger M_x|\psi\rangle}{\langle\psi|\psi\rangle} \quad (2.13)$$

Finally, any interaction with the system that leads to the measurement of a specific outcome x transforms any state vector $|\psi\rangle$ before the measurement to a new state vector $|\psi'\rangle$ after the measurement according to

$$|\psi'\rangle = M_x |\psi\rangle \quad (2.14)$$

that depends on the outcome measured.

The probabilities of eq. (2.13) are well defined since

$$p(x) = \frac{\|M_x|\psi\rangle\|^2}{\|\psi\rangle\|^2} \geq 0 \quad \forall x \quad (2.15)$$

$$\sum_x p(x) = \frac{\langle\psi|\sum_x M_x^\dagger M_x|\psi\rangle}{\langle\psi|\psi\rangle} = 1 \quad (2.16)$$

We thus have that for any fixed measurement every state vector defines a probability distribution on the outcomes. In this sense, quantum states can be thought of as a generalization of probability distributions. (Example) (Composed systems?)

2.1.2 PVM and observables

There is a special class of quantum measurements we will be interested in: projection-valued measurements (PVM). PVMs are quantum measurements where the measurement operators $\{P_x\}$ are required to be orthogonal projectors, i.e.

$$P_x^\dagger = P_x \quad \forall x \quad (2.17)$$

$$P_x^2 = P_x \quad \forall x \quad (2.18)$$

and to form a complete set of orthogonal projectors

$$P_x P_y = \delta_{xy} P_x \quad \forall x, y \quad (2.19)$$

$$\sum_x P_x = \mathbb{I} \quad (2.20)$$

To understand these definitions we first state an important theorem from linear algebra about self-adjoint operators, a complete treatment and proof can be found in ??

Theorem 6 (Spectral theorem). *Let \mathcal{H} be a finite-dimensional complex Hilbert space and A a self-adjoint operator on \mathcal{H} . Then there exists an orthonormal basis of eigenvectors of A with real eigenvalues.*

This means that for any self-adjoint operator, we may decompose the Hilbert space in orthogonal linear subspaces of eigenvectors with the same eigenvalue, the *eigenspaces*.

For any orthogonal projector P , we can show that

$$\begin{aligned} P |\psi\rangle = \lambda |\psi\rangle &\implies P^2 |\psi\rangle = \lambda^2 |\psi\rangle = P |\psi\rangle = \lambda |\psi\rangle \\ &\implies \lambda^2 = \lambda \\ &\implies \lambda = 0, 1 \end{aligned} \quad (2.21)$$

and, given an orthonormal base of eigenvectors $\{|e_i\rangle\}$ we may partition it into the two subsets $\{|e_j^{(0)}\rangle\}$ and $\{|e_k^{(1)}\rangle\}$ respectively of eigenvectors with eigenvalue 0 and 1. Then by expressing a generic vector as a linear combination of this base, we have that

$$|\psi\rangle = \sum_i \langle e_i | \psi \rangle |e_i\rangle \implies P |\psi\rangle = \sum_k \langle e_k^{(1)} | \psi \rangle |e_k^{(1)}\rangle \quad (2.22)$$

so that any orthogonal projector "orthogonally projects" vectors to its eigenspace with eigenvalue 1, which is thus also its image.

We can now recognize that a set of orthonormal projectors is complete when the spaces on which they project are orthogonal (eq. (2.19)) and add up to all the Hilbert space (eq. (2.20)). With this setting, it can be easily shown that the following corollary holds

Corollary 6.1. *Any self-adjoint operator A with eigenvalues $\{\lambda_i\}$ may be expressed in terms of the projectors P_i of its eigenspaces as*

$$A = \sum_i \lambda_i P_i \quad (2.23)$$

We thus have that, intuitively, PVMs are defined by decomposing the Hilbert space into orthogonal subspaces and then assigning a certain outcome to each one. In fact, if $|\psi_x\rangle$ is in the projected space of P_x the probability distribution of the outcomes will be

$$p(y) = \frac{\langle \psi_x | P_y^\dagger P_y | \psi_x \rangle}{\langle \psi_x | \psi_x \rangle} = \delta_{xy} \quad (2.24)$$

and also, after the measurement, any state vector is projected onto the subspace of the outcome

$$|\psi'\rangle = P_x |\psi\rangle \quad (2.25)$$

With this interpretation, it's easy to recognize an interesting property of PVMs: repeating the same PVM multiple times while the Hilbert space is evolving with the identity operator (i.e. is not changing) always leads to identical results. In fact, after the first measurement, if the measured outcome was x , we will have that

$$|\psi'_{(0)}\rangle = P_x |\psi\rangle \quad (2.26)$$

for any initial vector state $|\psi\rangle$. Then, repeating the same measurement, we will have the probability distribution

$$p(y) = \frac{\langle \psi | P_x^\dagger P_y^\dagger P_y P_x | \psi \rangle}{\langle \psi | P_x^\dagger P_x | \psi \rangle} = \delta_{xy} \quad (2.27)$$

and so we will get with certainty the same result. After the measurement, we will have the state vector

$$|\psi'_{(1)}\rangle = P_x P_x |\psi\rangle = P_x |\psi\rangle \quad (2.28)$$

and we can reiterate the same argument for the following measurements.

When considering well-defined measurable properties of a system, we may require repeated measurements to always give the same outcome if the system remains unchanged in between. From this intuitive concept follows the definition of an *observable* as a PVM with real-valued outcomes. Then, from Corollary 6.1, there is a 1-1 relationship between observables and self-adjoint operators through eq. (2.23), allowing us to identify any observable with its self-adjoint operator

$$\{P_i, \lambda_i\} \leftrightarrow A = \sum_i \lambda_i P_i \quad (2.29)$$

The expectation value of the observable when the system is in a state described by a state vector $|\psi\rangle$ is then given by

$$E[A] = \sum_i \lambda_i p_i = \sum_i \lambda_i \frac{\langle \psi | P_i^\dagger P_i | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\langle \psi | A | \psi \rangle}{\langle \psi | \psi \rangle} \quad (2.30)$$

2.1.3 Quantum states

Following postulates 1 to 3, if we are given a state vector for the system, we can determine how it will evolve under unitary evolutions, the probability distributions of the outcomes of quantum measurements, and the vector state we will obtain after those measurements, depending on the outcomes. We may then ask if there are multiple state vectors that yield the same probabilities for any measurement and continue to do so after any unitary evolution or measurement. Such two vectors would be completely equivalent in their predictions, and we can regard them as describing the same quantum state. This leads to an equivalence relation between state vectors, so that quantum states are the equivalence classes.

We start by requiring that two equivalent state vectors have the same probability distributions for any measurement. This means that

$$|\psi\rangle \sim |\psi'\rangle \iff \frac{\langle \psi | M_x^\dagger M_x | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\langle \psi' | M_x^\dagger M_x | \psi' \rangle}{\langle \psi' | \psi' \rangle} \quad \forall x \quad (2.31)$$

for any quantum measurement $\{(M_x, x)\}$. It's easy to see that a sufficient condition is

$$\exists c \neq 0 \in \mathbb{C} : |\psi'\rangle = c |\psi\rangle \quad (2.32)$$

since for any operator M_x

$$\frac{\langle \psi' | M_x^\dagger M_x | \psi' \rangle}{\langle \psi' | \psi' \rangle} = \frac{\|c\|^2}{\|c\|^2} \cdot \frac{\langle \psi | M_x^\dagger M_x | \psi \rangle}{\langle \psi | \psi \rangle} \quad (2.33)$$

Then by choosing $M_x = |\psi\rangle\langle\psi|$, it's easy to show that eq. (2.32) is also a necessary condition.

We also note that for any operator A

$$A(c|\psi\rangle) = cA|\psi\rangle \quad \forall c \in \mathbb{C} \quad (2.34)$$

so that any operator has a well-defined action on equivalence classes. In particular, this holds for unitary operators and measurement operators, and so two state vectors in the same equivalence class will remain in the equivalence class after any unitary evolution or measurement.

We have thus proved that quantum states are the equivalence classes of

$$|\psi\rangle \sim |\psi'\rangle \iff \exists c \neq 0 \in \mathbb{C} : |\psi'\rangle = c|\psi\rangle \quad (2.35)$$

Then for any state vector $|\psi\rangle$, its quantum state is

$$[|\psi\rangle]_\sim = \{c|\psi\rangle \mid c \neq 0 \in \mathbb{C}\} \quad (2.36)$$

and so the space of quantum states is isomorphic to the set of 1-dim linear subspaces of \mathcal{H} , also called *complex lines*. This set is known as the *projective space* of \mathcal{H} and is denoted as $\mathbf{P}\mathcal{H}$. When $\mathcal{H} = \mathbb{C}^N$, the projective space is known as the n -dimensional complex projective space \mathbb{CP}^n , where $n = N - 1$. Then eq. (2.36) leads to the intrinsic definition of a surjective *projection map*

$$\pi : \mathcal{H}_0 \rightarrow \mathbf{P}\mathcal{H} \quad \pi(|\psi\rangle) = [|\psi\rangle]_\sim \quad (2.37)$$

that maps every state vector to its quantum state.

2.1.4 Density operators

We have seen that quantum states are the 1-dim linear subspaces of \mathcal{H} , so we can identify them with some orthogonal projectors. The only additional requirement is that the projected spaces must be 1-dim, i.e., they must be rank-1 orthogonal projectors. From eq. (2.22), we know that for any rank-1 projector, there must exist a normalized vector $|\tilde{e}\rangle$ such that

$$P_{\tilde{e}}|\psi\rangle = \langle\tilde{e}|\psi\rangle|\tilde{e}\rangle \quad \forall|\psi\rangle \in \mathcal{H}_0 \quad (2.38)$$

and so from eq. (2.4), we have

$$P_{\tilde{e}} = |\tilde{e}\rangle\langle\tilde{e}| \quad (2.39)$$

In the context of quantum mechanics, the set of rank-1 orthogonal projectors is known as the set of *pure density operators* and is denoted as \mathfrak{D} .

We now define a candidate for the *projection map* to pure density operators

$$\Pi : \mathcal{H}_0 \rightarrow \mathfrak{D} \quad \Pi(|\psi\rangle) = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle} := \rho_\psi \quad (2.40)$$

This map is well-defined since for every $|\psi\rangle$ we have that ρ_ψ is self-adjoint, idempotent and

$$\rho_\psi = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle} = \left(\frac{1}{\|\psi\|} |\psi\rangle \right) \left(\langle\psi| \frac{1}{\|\psi\|} \right) \quad (2.41)$$

so that it is of rank 1. This map also has a well-defined action on quantum states since

$$\frac{(c|\psi\rangle)(\langle\psi|c^*)}{(\langle\psi|c^*)(c|\psi\rangle)} = \frac{|\psi\rangle\langle\psi|}{\langle\psi|\psi\rangle} \quad \forall c \neq 0 \in \mathbb{C} \quad (2.42)$$

so that it defines the map

$$f : \mathbf{PH} \rightarrow \mathfrak{D} \quad f([\psi]_\sim) = \Pi(|\psi\rangle) = \rho_\psi \quad (2.43)$$

Finally, we can prove that f is invertible, allowing us to identify quantum states with density matrices. The map is 1-1 since given any two state vectors $|\psi\rangle$ and $|\phi\rangle$, we have that

$$\begin{aligned} \rho_\psi = \rho_\phi &\implies \rho_\psi \rho_\phi = \rho_\phi \\ &\implies \frac{|\psi\rangle\langle\psi|\phi\rangle\langle\phi|}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle} = \frac{|\phi\rangle\langle\phi|}{\langle\phi|\phi\rangle} \\ &\implies \frac{\langle\psi|\phi\rangle}{\langle\psi|\psi\rangle} |\psi\rangle = |\phi\rangle \\ &\implies |\psi\rangle \sim |\phi\rangle \end{aligned}$$

Then, since any rank-1 orthogonal projector can be expressed as in eq. (2.39), we also have that f is surjective. We have thus proved that

$$\mathbf{PH} \sim \mathfrak{D} \quad (2.44)$$

through the intrinsic map f , so we may identify quantum states with pure density operators, and the two projection maps π and Π are the same under this identification.

We now want to restate postulates 1 to 3 in terms of pure density matrices. To do so, we will need the concept of the trace of an operator. For a finite-dimensional Hilbert space, we may define the trace of an operator A as the linear functional

$$\text{tr}(A) = \sum_i \langle e_i | A | e_i \rangle, \quad \{|e_i\rangle\} \text{ o.n.basis} \quad (2.45)$$

where it can be shown that the expression on the right is independent of the choice of the orthonormal basis. The trace also has the following cyclic property

$$\mathrm{tr}(AB) = \mathrm{tr}(BA) \quad (2.46)$$

so that for any number of operators A_1, \dots, A_n , we have

$$\mathrm{tr}(A_1 \dots A_n) = \mathrm{tr}(A_n A_1 \dots A_{n-1}) \quad (2.47)$$

We also have the following result from linear algebra

Theorem 7. *Let \mathcal{H} be a finite-dimensional complex Hilbert space and A any operator on \mathcal{H} . Then*

$$\mathrm{tr}(A) = \sum_i \lambda_i \quad (2.48)$$

where $\{\lambda_i\}$ are the eigenvalues of A repeated according to their algebraic multiplicity.

One immediate consequence of this theorem is that an orthogonal projector is of rank 1 if and only if it has a unitary trace. Then we may equivalently define pure density operators in the following, more common, way

$$\mathfrak{D} := \{\rho \in \mathfrak{L} \mid \rho^\dagger = \rho, \rho^2 = \rho, \mathrm{tr}(\rho) = 1\} \quad (2.49)$$

We can now reformulate postulate 1 as

Postulate 1 (Quantum states). *The set of all the possible quantum states of a system is the projective space of the system's Hilbert space. Then every quantum state is uniquely determined by a pure density operator $\rho \in \mathfrak{D}$*

Given a unitary evolution of the system, we can compute the evolution of any density operator as follows. Any initial quantum state can be expressed as

$$\rho(t_1) = |\psi(t_1)\rangle \langle \psi(t_1)| \quad (2.50)$$

where $|\psi(t_1)\rangle$ is a normalized state vector, then for any unitary evolution $U(t_1, t_2)$ we have that $|\psi(t_2)\rangle$ remains normalized and

$$\begin{aligned} \rho(t_2) &= |\psi(t_2)\rangle \langle \psi(t_2)| \\ &= U(t_1, t_2) |\psi(t_1)\rangle \langle \psi(t_1)| U^\dagger(t_1, t_2) \\ &= U(t_1, t_2) \rho(t_1) U^\dagger(t_1, t_2) \end{aligned}$$

so that we can reformulate postulate 2 as

Postulate 2 (Unitary evolution). *The Hilbert space of a closed system evolves only through unitary transformations. Coherently, the time evolution of any pure density operator $\rho(t)$ describing the quantum state of the system is given by*

$$\rho(t_2) = U(t_1, t_2)\rho(t_1)U^\dagger(t_1, t_2) \quad (2.51)$$

where $U(t_1, t_2)$ is a unitary operator that only depends on t_1 and t_2 .

Given any quantum measurement $\{(M_x, x)\}$ and a quantum state ρ we have that

$$\rho = |\psi\rangle\langle\psi| \implies p(x) = \langle\psi|M_x^\dagger M_x|\psi\rangle = \|M_x|\psi\rangle\|^2 \quad (2.52)$$

for some normalized state vector $|\psi\rangle$. Then by choosing

$$\frac{M_x|\psi\rangle}{\|M_x|\psi\rangle\|} \quad (2.53)$$

as the first element of an orthonormal basis, we have that

$$\text{tr}(M_x \rho M_x^\dagger) = \frac{\langle\psi|M_x^\dagger M_x|\psi\rangle \langle\psi|M_x M_x^\dagger|\psi\rangle}{\|M_x|\psi\rangle\|^2} = \|M_x|\psi\rangle\|^2 \quad (2.54)$$

so that we can reformulate postulate 3 as

Postulate 3 (Quantum measurements). *Quantum measurements are described by a collection of pairs $\{(M_x, x)\}$ of measurement operators M_x and outcomes x such that the following completeness equation is satisfied*

$$\sum_x M_x^\dagger M_x = \mathbb{I} \quad (2.55)$$

where \mathbb{I} is the identity operator. Then, given a system in a quantum state described by a pure density operator ρ , the probability distribution of the outcomes is

$$p(x) = \text{tr}(M_x \rho M_x^\dagger) \quad (2.56)$$

Finally, any interaction with the system that leads to the measurement of a specific outcome x transforms the quantum state ρ before the measurement to a new quantum state ρ' after the measurement according to

$$\rho' = \frac{M_x \rho M_x^\dagger}{\text{tr}(M_x \rho M_x^\dagger)} \quad (2.57)$$

that depends on the outcome measured.

At last, we have that the expectation value of an observable A when the system is in a quantum state described by a pure density operator ρ is

$$\text{E}[A] = \sum_i \lambda_i \text{tr}(P_i \rho P_i^\dagger) = \text{tr}(\rho A) \quad (2.58)$$

2.2 The manifold of quantum states

2.2.1 Geometry of the Hilbert space

Complex manifolds are defined in analogy with real ones with the requirement of being locally isomorphic to \mathbb{C}^N for some N and with holomorphic transition functions between charts; the tangent space at each point is thus also isomorphic to \mathbb{C}^N . Every N -dimensional complex manifold is also a $2N$ -dimensional real manifold where every complex coordinate basis $\{e_1, \dots, e_N\}$ corresponds to a real coordinate basis $\{e_1, \dots, e_N, ie_1, \dots, ie_N\}$, this manifold is called the *realification* of the original complex one.

Finite-dimensional Hilbert spaces are isomorphic to \mathbb{C}^N for some N and so they are also trivial complex manifolds. Moreover, there is a canonical isomorphism between the tangent space at each point and the Hilbert space itself

$$T_{|\psi\rangle}\mathcal{H} \sim \mathcal{H} \quad \text{with} \quad \frac{\partial}{\partial\theta} \leftrightarrow \left. \frac{d|\psi(\theta)\rangle}{d\theta} \right|_{\theta=0} =: |d\psi_\theta\rangle \quad (2.59)$$

so that the inner product on the Hilbert space defines an inner product on the complex tangent space of every point, i.e., it defines a complex $\binom{0}{2}$ tensor field on the manifold. Hilbert spaces also are metric spaces with respect to the distance induced by the inner product

$$d(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\| = \sqrt{\langle \mathbf{X} - \mathbf{Y}, \mathbf{X} - \mathbf{Y} \rangle} \quad \forall \mathbf{X}, \mathbf{Y} \in \mathcal{H} \quad (2.60)$$

For finite-dimensional Hilbert spaces, the inner product is a hermitian form that may be expressed as

$$\langle \mathbf{X}, \mathbf{Y} \rangle = h(\mathbf{X}, \mathbf{Y}) = g(\mathbf{X}, \mathbf{Y}) + iw(\mathbf{X}, \mathbf{Y}) \quad (2.61)$$

where g is a real-valued symmetric bilinear form and w is a real-valued antisymmetric bilinear form. Then, the distance may be expressed as

$$d(\mathbf{X}, \mathbf{Y}) = \sqrt{g(\mathbf{X} - \mathbf{Y}, \mathbf{X} - \mathbf{Y})} \quad (2.62)$$

so that the real part of the inner product also defines a metric tensor on the Hilbert space as explained in section 1.2.2, endowing it with a Riemannian structure.

We thus have that the space of state vectors \mathcal{H}_0 is also a complex manifold with the Riemannian structure

$$g_{|\psi\rangle}(|d\psi_1\rangle, |d\psi_2\rangle) = \text{Re} [\langle d\psi_1 | d\psi_2 \rangle] \quad (2.63)$$

This metric tensor is intrinsically derived from the structure of the Hilbert space and so it is invariant under unitary transformations. The set $U(N)$ of

unitary transformations of \mathbb{C}^N is a Lie group and so the Lie algebra of its left-invariant vector fields are the killing vector fields of the Hilbert space. The integral curves of these vector fields also preserve the norm of the points in the Hilbert space, and it can be shown that in each point their tangent vectors form a real $(2N-1)$ -dimensional linear subspace of the tangent space. From this follows that the integral curves of the killing vector fields mesh to form a foliation of the realified Hilbert space where each leaf is the set of state vectors with a fixed norm and is thus isomorphic to S^{2N-1} . Then each leaf is a $(2N-1)$ -dimensional submanifold of the realified Hilbert space and inherits a Riemannian structure.

2.2.2 Fiber bundle structure of the Hilbert space

We are now interested in the geometry of the space of quantum states. For start we have that, with a sound choice of coordinates, one can easily verify that \mathbb{CP}^N is also a complex manifold. (proof?) . Then tangent vectors of \mathbf{PH} can be intrinsically mapped to operators of \mathcal{H} as follows

$$\frac{\partial}{\partial \theta} \in T_\rho \mathfrak{D} \mapsto d\rho_\theta \quad \text{with} \quad d\rho_\theta(\mathbf{X}) = \left. \frac{d}{d\theta} [\rho(\theta)(\mathbf{X})] \right|_{\theta=0} \quad \forall \mathbf{X} \in \mathcal{H} \quad (2.64)$$

where from the linearity of differentiation follows that $d\rho_\theta$ is a well-defined linear operator on \mathcal{H} .

Now we may ask ourselves if there is a natural way to induce a metric on \mathbf{PH} from the one of \mathcal{H} . Such an intrinsic metric would lead to a natural notion of distance between quantum states. To do this we will start by investigating the relationship between \mathcal{H} and \mathbf{PH} as complex manifolds.

Firstly let us now recall the intrinsic projection map of eq. (2.37)

$$\pi : \mathcal{H}_0 \rightarrow \mathbf{PH} \quad |\psi\rangle \mapsto [|\psi\rangle]_\sim = \{c|\psi\rangle \mid \forall c \neq 0 \in \mathbb{C}\}$$

it can be shown that it is a smooth surjective map between the two manifolds so that we may regard \mathcal{H}_0 as a fiber bundle with base space \mathbf{PH} . The fibers of the fiber bundle are the orbits of the action on \mathcal{H}_0 of the abelian Lie group

$$\mathbf{C}_0 = \{c \cdot \mathbb{I} \mid \forall c \neq 0 \in \mathbb{C}\} \quad (2.65)$$

Since this group acts smoothly and transitively on the fibers, we have that the fibers are isomorphic to \mathbf{C}_0 , and the fiber bundle is a principal fiber bundle.

One way to induce a metric on \mathbf{PH} would be to find a natural embedding of it into \mathcal{H}_0 . It is a known result from the theory of principal fiber bundles

that if such an embedding existed, then the fiber bundle would be the trivial $\mathcal{H}_0 = \mathbf{PH} \times \mathbf{C}_0$. Since this is not the case, there is no natural embedding, i.e., there is no way to choose a representative state vector for every quantum state in terms of the structure of the Hilbert space only.

2.2.3 The Fubini-Study metric

Now that we have described the fiber bundle structure that links \mathcal{H}_0 and \mathbf{PH} , we introduce two concepts that will be instrumental in our aim of inducing a metric on the space of quantum states: the push-forward of vectors and the pull-back of 1-forms.

Given a fiber bundle X with base space B , the projection map $\pi : X \rightarrow B$ allows us to locally project any smooth curve on the fiber bundle to a smooth curve on the base space. From this follows that we have a natural map

$$d\pi_x : T_x X \rightarrow T_{\pi(x)} B \quad \forall x \in X \quad (2.66)$$

such that given any smooth curve $\gamma(t)$ with $\gamma(0) = x$, then

$$d\pi_x(\gamma'(0)) = \left. \frac{d}{dt} [\pi(\gamma(t))] \right|_{t=0} \quad (2.67)$$

where we have used the identification of tangent vectors with equivalence classes of curves. From the linearity of the differentiation follows that $d\pi$ is also linear. Then the push-forward is the map between the tangent bundles of the fiber bundle and the base space such that

$$d\pi : TX \rightarrow TB \quad \text{with} \quad d\pi(x, v) = (\pi(x), d\pi_x(v)) \quad (2.68)$$

The pull-back of 1-forms is the dual map of the push-forward. Given any 1-form $\alpha \in T_{\pi(x)}^* B$, we define its pull-back on $T_x^* X$ as the 1-form $\alpha' \in T_x^* X$ such that

$$\alpha'(v) = \alpha(d\pi_x(v)) \quad \forall v \in T_x X \quad (2.69)$$

Then we have that given a 1-form in the cotangent space of some point of the base space, we can pull it back to the cotangent spaces of the entire corresponding fiber of the fiber bundle. From this follows that the pull-back is defined as a map from entire fields of 1-forms in the base space to entire fields of 1-forms in the fiber bundle. This is in contrast with the push forward since for vector fields to be pushed forward, it would be necessary that every vector on the same fiber was pushed to the same vector on the corresponding element of the base space. Finally, using the tensor product, we can similarly define the push-forward of $(p)_0$ tensors and the pull-back of $(0)_q$ tensor fields.

The structure we want to project onto \mathbf{PH} is the $\binom{0}{2}$ metric tensor field, and so we cannot simply push it forward. The fiber bundle we are interested in, \mathcal{H}_0 , is also a Riemannian principal fiber bundle. As we will see, this allows us to uniquely *lift* vector fields from the base space to the fiber bundle.

Given a fiber bundle X as before, we have that at every point $x \in X$, the curves through x that remain on the same fiber are projected to single points on the base manifold so that the push-forward of their tangent vectors must be the null vector. We define the vertical subspace of $T_x X$ as the kernel of the push-forward

$$V_x := \ker d\pi_x \subseteq T_x X \quad (2.70)$$

Then a vector field is vertical if its value at every point is in the vertical subspace. For a general fiber bundle, there may not be any vertical vector fields since the vertical subspaces may not be a smooth subset of its tangent bundle. If X is also a principal fiber bundle with structure group G , we have that the fibers are given by the mesh of left-invariant vector fields of G . Then the vertical vector fields are precisely the left-invariant vector fields of G , and the vertical subspace is isomorphic to the Lie algebra of G at every point

$$V_x \sim \mathfrak{g}_G \quad (2.71)$$

We may also choose at every point $x \in X$ a horizontal subspace O_x , i.e., a complementary subspace to the vertical one so that

$$T_x X = V_x \oplus O_x \quad (2.72)$$

In general, there is no intrinsic way to choose a horizontal subspace at every point, but if X is also a Riemannian manifold, then we can define the horizontal subspaces as the orthogonal complements of the vertical ones

$$O_x := V_x^\perp \quad (2.73)$$

and horizontal vector fields are defined in analogy with the vertical ones. We thus have a well-defined notion of vertical motion, tangent to the fibers, and of horizontal motion, orthogonal to the fibers.

Intuitively, we can now lift any curve in the base space to the fiber bundle by starting from a point on the fiber and then moving orthogonally to it. To precisely define this concept, let us consider the restriction to O_x of the push-forward at every point

$$d\pi'_x \equiv d\pi_x|_{O_x} : O_x \subseteq T_x X \rightarrow T_{\pi(x)} B \quad (2.74)$$

It can be shown that from the smoothness of the metric tensor field follows that $d\pi'_x$ defines a smooth map between the tangent bundles of the fiber

bundle and the base space. Then from our previous considerations follows that

$$\ker d\pi'_x = \{\mathbf{0}\} \quad \forall x \in X \quad (2.75)$$

and thus $d\pi'_x$ defines an isomorphism between O_x and $T_{\pi(x)}B$ at every point. We can then uniquely define the *horizontal lift* $v_l(x) \in T_x X$ to the fiber bundle of any vector field $v(y) \in T_y B$ on the base space as

$$v_l(x) = d\pi_y'^{-1}(v(y)), \quad y = \pi(x) \quad \forall x \in X \quad (2.76)$$

that is the unique horizontal vector field on the fiber bundle that projects to $v(y)$.

We can now try to induce a metric tensor field on the base space by identifying its action on vector fields with the action of the metric tensor of the fiber bundle on the horizontal lift of the vector fields of the base space, i.e., given $dy_1, dy_2 \in T_y B$

$$g'_y(dy_1, dy_2) = g_x(d\pi_y'^{-1}(dy_1), d\pi_y'^{-1}(dy_2)) \quad (2.77)$$

where $y = \pi(x)$. This is possible only if the scalar field resulting from the inner product of the horizontal lifts is constant on the fibers; only in this case, we can then push forward the scalar field to the base space and thus define the action of the induced metric tensor field on the original vector fields. Requiring the inner product of the horizontal lifts to be constant on the fibers is equivalent to requiring it to be constant under the action of the structure group G . By definition, horizontal lifts are invariant under the action of G , so the only requirement is that the metric tensor field must be invariant, i.e., the left-invariant vector fields must also be killing vector fields.

We have thus shown that we can induce a metric tensor field on the base space of a Riemannian principal fiber bundle if the elements of the structure group are isometries of the fiber bundle. As we have seen in section 2.2.1, the set of isometries of $\mathcal{H}_0 = \mathbb{C}_0^N$ is the unitary group $U(N)$, while its structure group is the abelian group \mathbf{C}_0 . Then not every element of the structure group is an isometry since

$$\mathbf{C}_0 = \mathbb{R}^+ \times U(1) \quad c = \|c\| \cdot e^{i\theta} \leftrightarrow (\|c\|, e^{i\theta}) \quad (2.78)$$

To proceed, we must first fibrate with \mathbb{R}^+ and induce a metric on the base of this fibration. This is easily done since we can intrinsically define the embedding of the projection as the normalized state vectors

$$|\psi\rangle \mapsto \frac{|\psi\rangle}{\|\psi\|} = \frac{|\psi\rangle}{\sqrt{\langle\psi|\psi\rangle}} \in \mathcal{H}_0 \quad \forall |\psi\rangle \in \mathcal{H}_0 \quad (2.79)$$

so that the base space is one of the leaves of the foliation induced by $U(N)$ on \mathcal{H}_0

$$S^{2N-1} = \mathbb{C}_0^N / \mathbb{R}^+ \quad (2.80)$$

as described in section 2.2.2, then it also inherits the Riemannian structure such that

$$g_{|\psi\rangle}(|d\psi_1\rangle, |d\psi_2\rangle) = \text{Re} [\langle d\psi_1 | d\psi_2 \rangle] \quad (2.81)$$

for any normalized state vector $|\psi\rangle$. Considering the restriction of the projection map of eq. (2.37) on the set of normalized state vectors, we can repeat the same arguments of section 2.2.2 and show that S^{2N-1} is a Riemannian principal fiber bundle with structure group $U(1)$ and base space

$$\mathbb{CP}^n = S^{2N-1} / U(1) \quad (2.82)$$

where $n = N - 1$. The structure group is then a one-parameter subgroup of the isometry group, so we can induce a metric tensor field on \mathbb{CP}^n as described in this section. This metric for \mathbb{CP}^n is known as the Fubini-Study metric g_{FS} and is thus intrinsically defined in terms of the structure of the Hilbert space only.

The pullback of the FS metric on state vectors

We first derive an explicit expression for the pullback of the Fubini-Study metric on \mathcal{H}_0 . From our definitions, we have that for any normalized state vector $|\psi\rangle$

$$g_{FS}(|d\psi_1\rangle, |d\psi_2\rangle) = g(d\pi_\rho'^{-1}(d\rho_1), d\pi_\rho'^{-1}(d\rho_2)) \quad (2.83)$$

where $\rho = \pi(|\psi\rangle)$ and $d\rho_i = d\pi_{|\psi\rangle}(|d\psi_i\rangle)$. We first notice that the linear map

$$d\pi_\rho'^{-1} \circ d\pi_{|\psi\rangle} : T_{|\psi\rangle} \mathcal{H} \rightarrow O_{|\psi\rangle} \quad (2.84)$$

must act as the identity on $O_{|\psi\rangle}$ and must have kernel $V_{|\psi\rangle}$, the same as $d\pi_{|\psi\rangle}$, so that we can recognize it as the orthogonal projector on $O_{|\psi\rangle}$. Since for every $|\psi\rangle$, the motion along its fiber is given by $c|\psi\rangle$ for some complex number c , then

$$V_{|\psi\rangle} = \{c|\psi\rangle \mid \forall c \in \mathbb{C}\} \quad (2.85)$$

and so the projector to the vertical space is ρ . The projector to the horizontal space is then $(\mathbb{I} - \rho)$ so that

$$d\pi_\rho'^{-1} \circ d\pi_{|\psi\rangle}(|d\psi\rangle) = (\mathbb{I} - \rho)|d\psi\rangle = |d\psi\rangle - \langle \psi | d\psi \rangle |\psi\rangle \quad (2.86)$$

and for normalized state vectors we have

$$\begin{aligned} g_{FS}(|d\psi_1\rangle, |d\psi_2\rangle) &= \text{Re}[(\langle d\psi_2 | - \langle \psi | \langle d\psi_2 | \psi \rangle)(|d\psi_1\rangle - \langle \psi | d\psi_1 \rangle |\psi\rangle)] \\ &= \text{Re}[\langle d\psi_2 | d\psi_1 \rangle - \langle d\psi_2 | \psi \rangle \langle \psi | d\psi_1 \rangle] \end{aligned} \quad (2.87)$$

Finally, it can be similarly shown that for any state vector $|\psi\rangle \in \mathcal{H}_0$ and two tangent vectors $|d\psi_1\rangle, |d\psi_2\rangle \in T_{|\psi\rangle}\mathcal{H}$ we have

$$g_{FS}(|d\psi_1\rangle, |d\psi_2\rangle) = \text{Re} \left[\frac{\langle d\psi_2|d\psi_1\rangle}{\langle\psi|\psi\rangle} - \frac{\langle d\psi_2|\psi\rangle \langle\psi|d\psi_1\rangle}{\langle\psi|\psi\rangle^2} \right] \quad (2.88)$$

The FS metric on pure density operators

We now want to derive an explicit expression for the Fubini-Study metric on \mathbb{CP}^n . As we have seen, the first fibration projects the state vectors to the normalized state vectors, which constitute a leaf of the foliation of submanifolds of \mathcal{H}_0 that are meshed by the integral curves of the left-invariant vector fields of $U(N)$. Then, at any point of S^{2N-1} , the tangent space is isomorphic to the Lie algebra of $U(N)$. This is the space of anti-hermitian operators since the action of any 1-parameter subgroup of $U(N)$ on a normalized state vector $|\psi\rangle$ can always be locally expressed as

$$U(\theta)|\psi\rangle = e^{iA\theta}|\psi\rangle \quad (2.89)$$

for some hermitian operator A . Then, for any tangent vector $|d\psi\rangle \in T_{|\psi\rangle}S^{2N-1}$, there must exist a unique hermitian operator A such that

$$|d\psi\rangle = \frac{d}{d\theta} (e^{iA\theta}|\psi\rangle) \Big|_{\theta=0} = iA|\psi\rangle \quad (2.90)$$

The push forward of this vector is then given by

$$\begin{aligned} d\rho &= \frac{d}{d\theta} (e^{iA\theta}\rho e^{-iA\theta}) \Big|_{\theta=0} = (iA\rho - \rho iA) \\ &= i[A, \rho] \end{aligned} \quad (2.91)$$

We can then express the pullback of the Fubini-Study metric on S^{2N-1} in terms of the push forward of the tangent vectors

$$\begin{aligned} g_{FS}(|d\psi_1\rangle, |d\psi_2\rangle) &= \text{Re} [\langle d\psi_2|d\psi_1\rangle - \langle d\psi_2|\psi\rangle \langle\psi|d\psi_1\rangle] \\ &= \text{Re} [\langle\psi|A_2 A_1|\psi\rangle - \langle\psi|A_2|\psi\rangle \langle\psi|A_1|\psi\rangle] \\ &= \text{Re} [\text{tr} (\rho (A_2 A_1 - A_2 \rho A_1))] \\ &= \frac{1}{2} \text{tr} (\rho A_2 A_1 + \rho A_1 A_2 - \rho A_2 \rho A_1 - \rho A_1 \rho A_2) \\ &= \frac{1}{2} \text{tr} ([A_1, \rho] [\rho, A_2]) = \frac{1}{2} \text{tr} (d\rho_1 d\rho_2) \end{aligned} \quad (2.92)$$

and so we find that the Fubini-Study metric on \mathbb{CP}^n is given by

$$g_{FS}(d\rho_1, d\rho_2) = \frac{1}{2} \text{tr} (d\rho_1 d\rho_2) \quad (2.93)$$

Chapter 3

Quantum information

3.1 Quantum information geometry

3.1.1 Observables and the tangent space

As we have seen in ??, quantum states can be interpreted as a generalization of probability distributions. We now want to further investigate the parallelisms between the geometry of quantum states and the one of probability distributions.

We start by recalling the representation of tangent vectors to quantum states as linear operators, that we defined in eq. (2.64)

$$\frac{\partial}{\partial \theta} \in T_\rho \mathfrak{D} \mapsto d\rho_\theta \quad \text{with} \quad d\rho_\theta(\mathbf{X}) = \left. \frac{d}{d\theta} [\rho(\theta)(\mathbf{X})] \right|_{\theta=0} \quad \forall \mathbf{X} \in \mathcal{H} \quad (3.1)$$

This is analogous to the mixture representation of tangent vectors to probability distributions we defined in eq. (1.10). We then recall from the last section of ?? that the mixture representation of any tangent vector $d\rho \in T_\rho \mathfrak{D}$ can be expressed as

$$d\rho = i[A, \rho] \quad (3.2)$$

for some Hermitian operator A . Then it is evident that $d\rho$ is also Hermitian and we may interpret it as an observable; further, we have

$$\text{tr}(d\rho_\theta) = \frac{d}{d\theta} [\text{tr}(\rho(\theta))] = 0 \quad (3.3)$$

that is equivalent to the expression ??.

To define an analogous to the exponential representation, we would have to divide the mixture representation by the probability distribution itself but

this is not possible since we are dealing with operators. Instead, we can use the implicit expression of eq. (1.22) and symmetrize it to obtain

$$d\rho = \frac{1}{2} [\rho L^{(\rho)} + L^{(\rho)} \rho] \quad (3.4)$$

where $L^{(\rho)}$ would serve as the analogous to the exponential representation of tangent vectors. To precisely define this operator we express $d\rho$ as in eq. (3.2), so that we obtain the following equation

$$i[A, \rho] = \frac{1}{2} \{\rho, L^{(\rho)}(A)\} \quad (3.5)$$

It can be proved that $L^{(\rho)}(A)$ is well-defined for any operator A when ρ is Hermitian and positive semi-definite^{??}. Then $L^{(\rho)}(A)$ is a linear map between operators and it is known as the *symmetric logarithmic derivative (SLD)* of A . Moreover, it can be shown that $L^{(\rho)}(A)$ is Hermitian if A is Hermitian, so that

$$L^{(\rho)} : \mathfrak{A} \rightarrow \mathfrak{A} \quad \text{with} \quad \mathfrak{A} := \{A \in \mathfrak{L}(\mathcal{H}) \mid A = A^\dagger\} \quad (3.6)$$

and thus the exponential representations of tangent vectors can also be interpreted as observables. In analogy with ?? we also have that its expectation value is zero since

$$\mathbb{E}[L^{(\rho)}] = \text{tr}(\rho L^{(\rho)}) = \frac{1}{2} \text{tr}(\rho L^{(\rho)} + L^{(\rho)} \rho) = \text{tr}(d\rho) = 0 \quad (3.7)$$

Finally, for pure density operators, we can find an explicit expression for the SLD as follows

$$d\rho = \frac{d}{d\theta} [\rho(\theta)]|_{\theta=0} = \frac{d}{d\theta} [\rho^2(\theta)]|_{\theta=0} = \rho d\rho + d\rho \rho \quad (3.8)$$

so that for any pure density operator $\rho \in \mathfrak{D}$

$$L_\theta^{(\rho)} = 2d\rho_\theta \quad \forall d\rho_\theta \in T_\rho \mathfrak{D} \quad (3.9)$$

and since the SLD is a linear map we have that

$$L^{(\rho)} : \mathfrak{A} \rightarrow T_\rho \mathfrak{D} \quad (3.10)$$

is surjective. We have thus shown that the mixture and exponential representations of tangent vectors to pure quantum states coincide up to a constant factor. (Talk about the space of the mixture representation) (Show that the analogy works if quantum states are the pure states?)

3.1.2 Quantum Fisher information

In light of the analogies between the space of quantum states and the space of probability distributions, we can now try to define the analogous of the Fisher information metric for quantum states. We start from the formulation of the Fisher information metric in terms of the exponential representation of tangent vectors, recalling ??

$$G_F(X, Y) = \mathbb{E}_p[X^{(e)}Y^{(e)}] \quad \forall X, Y \in T_p\mathcal{P} \quad (3.11)$$

that is defined in terms of the expectation value of the product of the exponential representations of tangent vectors. Then we can define the quantum Fisher information metric by analogy, symmetrizing the arguments

$$\begin{aligned} G_{QF}(d\rho_1, d\rho_2) &= \mathbb{E}_\rho \left[\frac{1}{2} \left(L_1^{(\rho)} L_2^{(\rho)} + L_2^{(\rho)} L_1^{(\rho)} \right) \right] \\ &= \text{tr} \left[\rho \frac{L_1^{(\rho)} L_2^{(\rho)} + L_2^{(\rho)} L_1^{(\rho)}}{2} \right] \quad \forall d\rho_1, d\rho_2 \in T_\rho\mathfrak{D} \end{aligned} \quad (3.12)$$

$$\begin{aligned} &= \text{tr} \left[\frac{1}{2} \left\{ \rho, L_1^{(\rho)} \right\} L_2^{(\rho)} \right] = \text{tr} \left[L_1^{(\rho)} \frac{1}{2} \left\{ \rho, L_2^{(\rho)} \right\} \right] \\ &= \text{tr} \left[d\rho_1 L_2^{(\rho)} \right] = \text{tr} \left[L_1^{(\rho)} d\rho_2 \right] \quad \forall d\rho_1, d\rho_2 \in T_\rho\mathfrak{D} \end{aligned} \quad (3.13)$$

so that from ?? follows that for pure quantum states

$$G_{QF}(d\rho_1, d\rho_2) = 2 \text{tr}[d\rho_1 d\rho_2] \quad \forall d\rho_1, d\rho_2 \in T_\rho\mathfrak{D} \quad (3.14)$$

$$= 4 g_{FS}(d\rho_1, d\rho_2) \quad \forall d\rho_1, d\rho_2 \in T_\rho\mathfrak{D} \quad (3.15)$$

This shows that the Fubini-Study metric we defined intrinsically for pure quantum states can be precisely interpreted as a quantum information metric.

3.1.3 Symmetric covariance

Our analogy between the geometry of quantum states and the one of probability distributions is now almost complete but there are still some details to be precised. For start, we would want to treat observables as we treated random variables but while every tangent vector can be associated with an observable with zero expectation value, the converse is not directly true since

$$T_\rho\mathfrak{D} \subset \mathfrak{A}_\rho^\perp := \{A \in \mathfrak{A}(\mathcal{H}) \mid \mathbb{E}_\rho[A] = 0\} \quad (3.16)$$

as can be easily checked by comparing the real dimensions of the two vector spaces. Moreover, since operators don't commute in general, we must define a

sound generalization of the covariance and as we will show these two problems are related.

For start, we require our generalized covariance to reduce to the variance of the observable when evaluated on the diagonal

$$\text{Cov}_\rho[A, A] = \text{V}_\rho[A] = \text{tr}(\rho(A - \text{E}_\rho[A]\mathbb{I})^2) \quad \forall A \in \mathfrak{A} \quad (3.17)$$

Then by requiring it to be symmetric in its arguments, we are naturally led to the definition

$$\text{Cov}_\rho[A, B] := \frac{1}{2} \text{tr}(\rho \{(A - \text{E}_\rho[A]\mathbb{I}), (B - \text{E}_\rho[B]\mathbb{I})\}) \quad \forall A, B \in \mathfrak{A} \quad (3.18)$$

Since $(A - \text{E}_\rho[A]\mathbb{I})$ always has zero expectation value, we can recognize that this expression also defines for every quantum state an inner product in the space of observables with zero expectation value

$$\ll A, B \gg_\rho := \text{E} \left[\frac{1}{2} \{A, B\} \right] \quad \forall A, B \in \mathfrak{A}^\perp \quad (3.19)$$

Then we can easily recognize that this inner product coincides with the quantum fisher metric for the exponential representations of tangent vectors of ??

$$\ll L_1^{(\rho)} L_2^{(\rho)} \gg_\rho = G_{QF}(d\rho_1, d\rho_2) \quad \forall d\rho_1, d\rho_2 \in T_\rho \mathfrak{D} \quad (3.20)$$

This is no coincidence and it can be shown that the definitions of the exponential representation and of the generalized covariance are deeply linked ????.

Let us now go back to the problem of associating every observable with zero expectation value to the exponential representation of a tangent vector. We notice that every observable $A \in \mathfrak{A}$ can be thought of as an element of the Lie algebra of $U(N)$ and thus we can associate it with a tangent vector through the linear map

$$A \mapsto \frac{d}{d\theta} (e^{iA\theta} \rho e^{-iA\theta}) \Big|_{\theta=0} = i[A, \rho] =: d\rho_A \quad (3.21)$$

Then the kernel of this map is the linear subspace of observables that commute with the pure density operator of the quantum state. We can easily interpret this considering the fact that if an observable commutes with the pure density operator of a quantum state, then the state vectors of that quantum state are eigenvectors of the observable. Let $\rho = |\psi\rangle\langle\psi|$ for some normalized state vector $|\psi\rangle$, then

$$[\rho, A] = 0 \implies \rho A = A \rho$$

$$\implies A|\psi\rangle = A\rho|\psi\rangle = \rho A|\psi\rangle = \langle\psi|A|\psi\rangle|\psi\rangle \quad (3.22)$$

$$\implies A\rho = \text{E}_\rho[A]\rho \quad (3.23)$$

and so the 1-parameter unitary subgroup of A will move state vectors parallel to the fiber, thus not changing the quantum state.

We can now consider the restriction of this map to the observables with zero expectation value. Through the matrix representation of operators we can easily compute the real dimensions of \mathfrak{A}^\perp and of the kernel of the map. These turn out to be respectively $N^2 - 1$ and $(N - 1)^2$ so that the image of the map must have real dimension $2N - 2$, equal to the dimension of the tangent space; this proves that the map is surjective. Then we can also show that the map maps $T_\rho \mathfrak{D}$ to itself since

$$\begin{aligned} [L^{(\rho)}, \rho] = 0 &\implies d\rho = L^{(\rho)}\rho = \rho L^{(\rho)} \\ &\implies (d\rho - L^{(\rho)})\rho = \rho(d\rho - L^{(\rho)}) \\ &\implies iA\rho - i\rho A\rho - L^{(\rho)}\rho = i\rho A - i\rho A\rho - \rho L^{(\rho)} \\ &\implies i[A, \rho] = [L^{(\rho)}, \rho] \\ &\implies d\rho = 0, L^{(\rho)} = 0 \end{aligned}$$

where we considered $d\rho = i[A, \rho]$. From this follows that we can decompose the vector space of observables with zero expectation value as

$$\mathfrak{A}^\perp = \mathfrak{A}_{C(\rho)}^\perp \oplus T_\rho \mathfrak{D} \quad \forall \rho \in \mathfrak{D} \quad (3.24)$$

where $\mathfrak{A}_{C(\rho)}^\perp$ is the set of zero expectation value observables that commute with ρ . Given any observable with zero expectation value we write

$$\begin{aligned} A = A_{C(\rho)} + A_\rho^{(e)} &\quad \forall A \in \mathfrak{A}^\perp \\ \text{with } A_{C(\rho)} \in \mathfrak{A}_{C(\rho)}^\perp, A_\rho^{(e)} \in T_\rho \mathfrak{D} \end{aligned} \quad (3.25)$$

and we thus have that the decomposition is unique.

Finally, we can show that the symmetric covariance we defined only depends on the non-commuting part of the decomposition

$$\begin{aligned} \ll A, B \gg_\rho &= \frac{1}{2} \text{tr} [\rho \{A, B\}] \\ &= \frac{1}{2} \text{tr} [\rho \{A_{C(\rho)} + A_\rho^{(e)}, B_{C(\rho)} + B_\rho^{(e)}\}] \\ &= \frac{1}{2} \text{tr} [\rho \{A_\rho^{(e)}, B_\rho^{(e)}\}] + \frac{1}{2} \text{tr} [\rho \{A_\rho^{(e)}, B_{C(\rho)}\}] + \\ &\quad + \frac{1}{2} \text{tr} [\rho \{A_{C(\rho)}, B_\rho^{(e)}\}] + \frac{1}{2} \text{tr} [\rho \{A_{C(\rho)}, B_{C(\rho)}\}] \\ &= \frac{1}{2} \text{tr} [\rho \{A_\rho^{(e)}, B_\rho^{(e)}\}] = \ll A_\rho^{(e)}, B_\rho^{(e)} \gg_\rho \end{aligned} \quad (3.26)$$

so that we can define the map

$$A \mapsto A_\rho \in T_\rho \mathbf{P}\mathcal{H} \quad \forall A \in \mathfrak{A}^\perp, \rho \in \mathbf{P}\mathcal{H} \quad (3.27)$$

where A_ρ is the tangent vector whose exponential representation is $A_\rho^{(e)}$. Then we have also proved that

$$\ll A, B \gg_\rho = G_{QF}(A_\rho, B_\rho) \quad (3.28)$$

and we have thus arrived at an exact analogy of ??.