Contents

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 $p \in \mathbb{R}^{\mathcal{X}} \coloneqq \{f \mid f \colon \mathcal{X} \to \mathbb{R}\}$ 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

$$\mathrm{E}_p[A] \coloneqq \sum_{x \in \mathcal{X}} p(x) A(x) \tag{1.2}$$

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

$$\mathrm{Cov}_p[A, B] \coloneqq \mathrm{E}_p[(A - \mathrm{E}_p[A])(B - \mathrm{E}_p[B])] \tag{1.3}$$

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

$$\mathrm{V}_p[A] \coloneqq \mathrm{Cov}_p[A, A] = \mathrm{E}_p[(A - \mathrm{E}_p[A])^2] \tag{1.4}$$

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 $\mathbb{R}^N$

$$f \leftrightarrow (f(x_1), \dots, f(x_N)) \tag{1.5}$$

then it's easy to recognize that the *space of probability distributions* is a convex subset of the affine subspace $\mathcal{A}_1 \coloneqq \{f \in \mathbb{R}^{\mathcal{X}} \mid \sum_{x\in\mathcal{X}} f(x) = 1\}$. 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

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

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 $\psi\colon \mathcal{S} \to \Xi \subseteq \mathbb{R}^n$, so that we may write

$$\mathcal{S} = \left\{ p_\xi \;\middle|\; \exists\, \xi = (\xi^{(1)}, \dots, \xi^{(n)}) \in \Xi\colon p_\xi = \psi^{-1}(\xi) \right\} \tag{1.7}$$

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 $\partial_i \coloneqq \frac{\partial}{\partial \xi^{(i)}}$. 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 $\psi' \colon \mathcal{S} \to \Xi' \subseteq \mathbb{R}^n$ that follows eq. (1.8), the composed function $\psi' \circ \psi^{-1} \colon \Xi \to \Xi'$ 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

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

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 $\mathcal{A}_0 \coloneqq \{A \in \mathbb{R}^{\mathcal{X}} \mid \sum_{x \in \mathcal{X}} A(x) = 0\}$. 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

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

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

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

since

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

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

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

is an isomorphism and

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

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

$$\mathcal{A}_p^{\perp} \coloneqq \{A \in \mathbb{R}^{\mathcal{X}} \mid p \cdot A = \mathrm{E}_p[A] = 0\} \tag{1.15}$$

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 $\mathbb{R}^{\mathcal{X}}$

$$p \mapsto \ln p \in \mathbb{R}^{\mathcal{X}} \tag{1.16}$$

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

$$X^{(e)}(x) \coloneqq X(\ln p(x)) = \frac{X(p(x))}{p(x)} \tag{1.17}$$

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

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

since

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

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

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

is an isomorphism and

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

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

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

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 $N_s \to \infty$

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

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

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

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*

$$P_{N_s}(\mathcal{E}) \sim e^{-N_s S(p_* \| q)} \qquad \text{for} \quad N_s \to \infty \tag{1.25}$$

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

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

$$S(p \parallel q) = 0 \iff p = q \tag{1.27}$$

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 $S(p \parallel q) \to \infty$ 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 $S(p \parallel q) \to \infty$ 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*

$$S(p \parallel q) \geq S(p \parallel p_*) + S(p_* \parallel q) \tag{1.28}$$

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

$$\mathrm{Exp}_p : T_pM \to M, \quad \mathrm{Exp}_p(\mathbf{v}) \coloneqq \gamma_{\mathbf{v}}(1) \tag{1.29}$$

where $\mathbf{v} \in T_p$ and $\gamma_{\mathbf{v}} : [0,1] \to M$ 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 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 $q \mapsto \mathbf{v}_q \in T_pM$ 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

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

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

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

Finally then, we get

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

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

Now chose a vector $\mathbf{dp} \in T_pM$ and let $q = \mathrm{Exp}_p(\mathbf{dp})$. 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 $\gamma_{\mathbf{dp}}$

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

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 $\Gamma^{(i)}_{\hat{\mathbf{dp}}}(l)$ in $l = 0$ we get

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

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

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

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

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

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

$$L^2_p(\{p^{(i)} + dp^i\}) = \frac{1}{2}\Big[\partial_i \partial_j L^2_p(q)\Big]_{q=p} dp^i dp^j + O(dp^3) \tag{1.37}$$

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

$$g^{(p)}_{ij} = \frac{1}{2}\Big[\partial_i \partial_j L^2_p(q)\Big]_{q=p} \tag{1.38}$$

so the metric tensor in p is proportional to the Hessian matrix of L^2_p 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 $D(\,\cdot \parallel \cdot\,) : M \times M \to [0, +\infty)$ a C^2 function (possibly asymmetric) satisfying

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

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)

$$D[\,\partial_i \parallel \cdot\,] \colon p \mapsto [\partial_i D(q \parallel q')]_{(q,q')=(p,p)}$$
$$D[\,\cdot \parallel \partial_i\,] \colon p \mapsto [\partial'_i D(q \parallel q')]_{(q,q')=(p,p)}$$

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

$$D[\,\partial_i \parallel \cdot\,] = D[\,\cdot \parallel \partial_i\,] \equiv 0 \tag{1.40}$$

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

$$D[\,\partial_i\partial_j \parallel \cdot\,] = D[\,\cdot \parallel \partial_i\partial_j\,] = -D[\,\partial_i \parallel \partial_j\,] =: g^{(D)}_{ij} \tag{1.41}$$

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

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

to second order, we get

$$D^{(R)}_p(q) = \frac{1}{2}g^{(D)}_{ij}d\xi^i d\xi^j + O(d\xi^2) \quad \text{and} \quad D^{(L)}_p(q) = \frac{1}{2}g^{(D)}_{ij}d\xi^i d\xi^j + O(d\xi^2)$$
$$\text{where } d\xi^i \coloneqq \xi^{(i)}_p - \xi^{(i)}_q$$

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

Finally, if $g^{(D)}_{ij}$ is positive definite we say that D is a *divergence*, then $\frac{1}{2}g^{(D)}_{ij}$ 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

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

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

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

And so we have

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

and equivalently

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

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

where we used the fact that

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

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, $G_F \coloneqq \{g_{ij}^{(S)}\}$ 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

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

while from eq. (1.45) we get

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

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

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

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

then it's easy to guess that the embedding

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

is the one whose representation of the tangent spaces

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

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

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

(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

$$\hat{\xi} = (\hat{\xi}^{(1)}, \dots, \hat{\xi}^{(n)}) \colon \mathcal{X} \to \Xi \subseteq \mathbb{R}^n \tag{1.54}$$

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

$$\mathrm{E}_{p_\xi}\left[\hat{\xi}\right] = \left(\mathrm{E}_{p_\xi}\left[\hat{\xi}^{(1)}\right], \dots, \mathrm{E}_{p_\xi}\left[\hat{\xi}^{(n)}\right]\right) = \xi \qquad \forall \xi \in \Xi \tag{1.55}$$

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 $V_\xi\left[\hat{\xi}\right] \coloneqq \left\{v_\xi^{ij}\right\}$ where

$$v_\xi^{ij} \coloneqq \mathrm{Cov}_{p_\xi}\left[\hat{\xi}^{(i)}, \hat{\xi}^{(j)}\right] = \mathrm{E}_{p_\xi}\left[\left(\hat{\xi}^{(i)}(x) - \xi^i\right)\left(\hat{\xi}^{(j)}(x) - \xi^j\right)\right] \tag{1.56}$$

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 $\mathrm{E}[A]$ 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

$$\mathrm{E}[A]\colon \mathcal{P} \to \mathbb{R} \qquad p \mapsto \mathrm{E}_p[A] \tag{1.57}$$

Since this is a function of the manifold $\mathcal{P}$, at every point p we may consider its differential $(d\mathrm{E}[A])_p$. 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

$$(d\mathrm{E}[A])_p(X) = X(\mathrm{E}[A]) \tag{1.58}$$

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

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

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*

$$(grad\mathrm{E}[A])_p^{(e)} = A - \mathrm{E}_p[A] \qquad \forall p \in \mathcal{P} \tag{1.60}$$

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

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

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

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

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

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

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

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

□

We also get the following

Corollary 3.1. *For any random variable A*

$$\mathrm{V}_p[A] = \|(d\mathrm{E}[A])_p\|_p^2 \tag{1.61}$$

Proof. Follows from theorem 3 noticing that

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

□

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

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

and so we may uniquely decompose

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

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

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

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

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*

$$\mathrm{V}_p[A] \geq \|(d\mathrm{E}[A]|_{\mathcal{S}})_p\|_p^2 \tag{1.65}$$

where the equality holds if and only if

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

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 $\mathrm{V}_\xi[\hat{\xi}]$ satisfies*

$$\mathrm{V}_\xi[\hat{\xi}] \geq G_F^{-1}(p_\xi) \tag{1.67}$$

in the sense that $\mathrm{V}_\xi[\hat{\xi}] - G_F^{-1}(p_\xi)$ 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 $\mathrm{E}_{p_\xi}[A] = c_i\xi^i$ and

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

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

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

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

$$\begin{aligned}\|(d\mathrm{E}[A]|_{\mathcal{S}})_p\|_p^2 &= (\partial_i \mathrm{E}[A])_p g^{ij}(p)(\partial_j \mathrm{E}[A])_p \\ &= c_i g^{ij}(p) c_j\end{aligned}$$

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 The manifold of quantum states

2.1.1 The postulates of quantum mechanics

Quantum mechanics is a fundamental theory that prescribes a mathematical framework to model the states of physical systems and their evolution, and that describes how to predict the results of our observations of them. We begin by stating its postulates loosely following the treatment of **??**.

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 physical meaning of this condition will be discussed later in the section.

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

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

Similarly, given any linear mapping A from $\mathcal{H}$ to itself, that from now on we will call *operator*, we can define its adjoint as the operator $\mathrm{A}^\dagger$ such that

$$\langle \mathrm{A}^\dagger\, \mathbf{X}, \mathbf{Y} \rangle = \langle \mathbf{X}, \mathrm{A}\, \mathbf{Y} \rangle \qquad \forall\, \mathbf{X}, \mathbf{Y} \in \mathcal{H} \tag{2.2}$$

it can be shown that this is always well-defined for finite-dimensional Hilbert spaces and that $(\mathrm{A}^\dagger)^\dagger = \mathrm{A}$.

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.

$$\langle \mathrm{U}\mathbf{X}, \mathrm{U}\mathbf{Y}\rangle = \langle \mathbf{X}, \mathbf{Y}\rangle \qquad \forall \mathbf{X}, \mathbf{Y} \in \mathcal{H} \tag{2.3}$$

From eq. (2.2) it's easy to show that

$$U \textit{ is unitary} \iff \mathrm{U}\mathrm{U}^\dagger = \mathbb{I} \textit{ i.e } \mathrm{U}^{-1} = \mathrm{U}^\dagger \tag{2.4}$$

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

$$\langle \mathrm{A}\,\mathbf{X}, \mathbf{Y}\rangle = \langle \mathbf{X}, \mathrm{A}\,\mathbf{Y}\rangle \qquad \forall\, \mathbf{X}, \mathbf{Y} \in \mathcal{H} \tag{2.5}$$

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

$$\langle \mathrm{A}\,\mathbf{X}, \mathbf{X}\rangle = \overline{\langle \mathbf{X}, \mathrm{A}\,\mathbf{X}\rangle} = \overline{\langle \mathrm{A}\,\mathbf{X}, \mathbf{X}\rangle} \in \mathbb{R} \qquad \forall\, \mathbf{X} \in \mathcal{H} \tag{2.6}$$

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*. The state vectors of the Hilbert space 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 $\langle\psi|$. Then, from eq. (2.1) follows that we may write unambiguously

$$\langle\psi|\phi\rangle \equiv \langle |\psi\rangle\,, |\phi\rangle\rangle = \big[\langle\psi|\big](|\phi\rangle) \tag{2.7}$$

where the operation being done may be equivalently interpreted as the dual functional of $|\psi\rangle$ acting on $|\phi\rangle$ or as the inner product of the two vectors. The image of a vector $|\psi\rangle$ under the action of an operator L will be written as $\mathrm{L}\,|\psi\rangle$, then from the definition of adjoint operators follows that its dual will be $\langle\psi|\,\mathrm{L}^\dagger$. Finally, given any self-adjoint operator A, from eq. (2.5) follows that we may write unambiguously

$$\langle\psi|A|\phi\rangle \equiv [\langle\psi|\,A]\,|\phi\rangle = \langle\psi|\,[A\,|\phi\rangle] \tag{2.8}$$

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*

$$|\psi(t_2)\rangle = \mathrm{U}(\mathrm{t}_1, \mathrm{t}_2)\,|\psi(t_1)\rangle \tag{2.9}$$

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* $\{(\mathrm{M}_x, x)\}$ *of* measurement operators M_x *and* outcomes x *such that the following* completeness equation *is satisfied*

$$\sum_x \mathrm{M}_x^\dagger\,\mathrm{M}_x = \mathbb{I} \tag{2.10}$$

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*

$$p(x) = \frac{\langle\psi|\mathrm{M}_x^\dagger\,\mathrm{M}_x|\psi\rangle}{\langle\psi|\psi\rangle} \tag{2.11}$$

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*

$$|\psi'\rangle = \mathrm{M}_x\,|\psi\rangle \tag{2.12}$$

that depends on the outcome measured.

The probabilities of eq. (2.11) are well defined since

$$p(x) = \frac{\|\mathrm{M}_x\,|\psi\rangle\|^2}{\||\psi\rangle\|^2} \geq 0 \qquad \forall x \tag{2.13}$$

$$\sum_x p(x) = \frac{\langle\psi|\sum_x \mathrm{M}_x^\dagger\,\mathrm{M}_x|\psi\rangle}{\langle\psi|\psi\rangle} = 1 \tag{2.14}$$

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 and tensor product)

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.

$$P_x^\dagger = P_x \qquad \forall x \tag{2.15}$$
$$P_x^2 = P_x \qquad \forall x \tag{2.16}$$

and to form a complete set of orthogonal projectors

$$P_x\, P_y = \delta_{xy} P_x \qquad \forall x, y \tag{2.17}$$
$$\sum_x P_x = \mathbb{I} \tag{2.18}$$

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

$$\begin{aligned} P\,|\psi\rangle = \lambda\,|\psi\rangle \implies & P^2\,|\psi\rangle = \lambda^2\,|\psi\rangle = P\,|\psi\rangle = \lambda\,|\psi\rangle \\ \implies & \lambda^2 = \lambda \\ \implies & \lambda = 0, 1 \end{aligned} \tag{2.19}$$

so that any orthogonal projector "projects" vectors to its eigenspace with eigenvalue 1. With this perspective, we recognize that a set of orthonormal projectors is complete when the spaces on which they project are orthogonal (eq. (2.17)) and add up to all the Hilbert space (eq. (2.18)).

Intuitively PVM are defined by decomposing the Hilbert space in 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

$$p(y) = \frac{\langle\psi_x|P_y^\dagger\, P_y|\psi_x\rangle}{\langle\psi_x|\psi_x\rangle} = \delta_{xy} \tag{2.20}$$

PVMs also have the interesting property that repeating the same measurement 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

$$|\psi'_{(0)}\rangle = P_x |\psi\rangle \tag{2.21}$$

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

$$p(y) = \frac{\langle\psi|P_x^\dagger P_y^\dagger P_y P_x|\psi\rangle}{\langle\psi|P_x^\dagger P_x|\psi\rangle} = \delta_{xy} \tag{2.22}$$

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

$$|\psi'_{(1)}\rangle = P_x P_x |\psi\rangle = P_x |\psi\rangle \tag{2.23}$$

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

When we think of well-defined measurable properties of a system we may require repeated measurements to always give the same outcome if in between them the system was unchanged. From this intuitive concept follows the definition of an *observable* as a PVM with real-valued outcomes λ_i. Then the following corollary allows us to link observables and self-adjoint operators

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

$$\mathrm{A} = \sum_i \lambda_i P_i \tag{2.24}$$

From this immidiately follows that letting $\{\lambda_i\}$ be the real outcomes and P_i be their PVM measurement there is a 1-1 relationship between observables and selfadjoint operators so that we may identify any observable with its selfadjoint operator. (expectation value?)

2.1.3 Quantum states

Following postulates 1 to 3 we have that given a state vector for the system we know how it will evolve under some unitary evolution, the probability distributions of the outcomes of quantum measurements we may do, and the vector state we will get after those measurements, depending on the outcomes. Then we may ask ourselves if there are multiple state vectors

that for any measurement yield the same probabilities, and that continue to do so after any unitary evolution or measurement. Such two vectors would be completely equivalent in their predictions and so we may regard them as describing the same quantum state. What we have just described is 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

$$|\psi\rangle \sim |\psi'\rangle \iff \frac{\langle\psi|\mathrm{M}_x^\dagger \mathrm{M}_x|\psi\rangle}{\langle\psi|\psi\rangle} = \frac{\langle\psi'|\mathrm{M}_x^\dagger \mathrm{M}_x|\psi'\rangle}{\langle\psi'|\psi'\rangle} \qquad \forall x \tag{2.25}$$

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

$$\exists\, c \in \mathbb{C} : \quad |\psi'\rangle = c\,|\psi\rangle \tag{2.26}$$

since for any operator M_x

$$\frac{\langle\psi'|\mathrm{M}_x^\dagger \mathrm{M}_x|\psi'\rangle}{\langle\psi'|\psi'\rangle} = \frac{\|c\|^2}{\|c\|^2} \cdot \frac{\langle\psi|\mathrm{M}_x^\dagger \mathrm{M}_x|\psi\rangle}{\langle\psi|\psi\rangle} \tag{2.27}$$

Then by choosing $\mathrm{M}_x = P_\psi$, the projector on the span of ψ, it's easy to show that eq. (2.26) is also a necessary condition.

We also note that for any operator A

$$A(c\,|\psi\rangle) = cA\,|\psi\rangle \quad \forall c \in \mathbb{C} \tag{2.28}$$

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 remain in the equivalence class after any unitary evolution or measurement.

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

$$|\psi\rangle \sim |\psi'\rangle \iff \exists\, c \in \mathbb{C} : \quad |\psi'\rangle = c\,|\psi\rangle \tag{2.29}$$

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

$$[|\psi\rangle]_\sim = \{c\,|\psi\rangle \ \forall c \in \mathbb{C}\} \setminus \{0\} \tag{2.30}$$

and so the space of quantum states is isomorphic to the set of 1-dim linear subspaces of $\mathcal{H}$, also called *complex lines*. When $\mathcal{H} = \mathbb{C}^N$ this set is known as the N-dimensional complex projective space $\mathbb{C}\mathbf{P}^N$ and it will be the object of our study.