7 The Bures distance and Uhlmann fidelity

In this section we study the Bures distance on the set of quantum states $\mathcal{E}(\mathcal{H})$. This distance is Riemannian and monotonous with respect to quantum operations. It is a simple function of the fidelity (96). Its metric coincides with the quantum Fisher information quantifying the best achievable precision in the parameter estimation problem discussed in Sec. 8.2. The material of this section (as well as of Sec. 8) is completely independent from that of sections 9 and 10, so it is possible at this point to proceed directly to Sec. 9. The reading of Secs. 7.1–7.4 is, however, recommended before going through Sec. 11 devoted to the geometrical measures of quantum correlations, where the Bures distance plays the key role. The section is organized as follows. Sec. 7.1 contains a short discussion on contractive (i.e., monotonous) distances. It is argued there that the distances induced by the $\|\cdot\|_p$-norm are not contractive save for $p = 1$. The definition and main properties of the Bures distance are given in Secs. 7.2–7.4. The Bures metric is determined in Sec.7.5. Finally, Sec. 7.8 contains the proof of an important result of Petz on the characterization of all Riemannian contractive metrics on $\mathcal{E}(\mathcal{H})$ for finite-dimensional Hilbert spaces $\mathcal{H}$.

7.1 Contractive and convex distances

In order to quantify how far are two states ρ and σ it is necessary to define a distance on the set $\mathcal{E}(\mathcal{H})$ of quantum states. One has a priori the choice between many distances. The most common ones are the L^p-distances defined by (2). In quantum information theory it seems, however, natural to impose the following requirement.

Definition 7.1.1. *A distance d on the sets of quantum states is contractive if for any finite Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$, any quantum operation $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$, and any ρ, $\sigma \in \mathcal{E}(\mathcal{H})$, it holds*

$$d(\mathcal{M}(\rho), \mathcal{M}(\sigma)) \leq d(\rho, \sigma) \ . \tag{169}$$

A contractive distance is in particular invariant under unitary conjugations, i.e.,

$$d\big(U\rho U^*, U\sigma U^*\big) = d(\rho, \sigma) \quad \text{if } U \text{ is unitary} \tag{170}$$

(in fact, $\rho \mapsto U\rho U^*$ is an invertible quantum operation on $\mathcal{B}(\mathcal{H})$). For such a distance, if a generalized measurement is performed on a system, two states are closer from each other after the measurement than before it, and if the system is subject to a unitary evolution the distance between the time-evolved states remains unchanged.

For $p > 1$, the distances d_p (in particular, the Hilbert-Schmidt distance d_2) are not contractive. A counter-example for two qubits is obtained [140] by taking $\mathcal{M}(\rho) = A_1\rho A_1^* + A_2\rho A_2^*$ with

$$A_1 = \sigma_+ \otimes 1 \quad , \quad A_2 = \sigma_+\sigma_- \otimes 1 \quad , \quad \rho = \frac{1}{2} \otimes \sigma_+\sigma_- \quad , \quad \sigma = \frac{1}{2} \otimes \sigma_-\sigma_+ \tag{171}$$

(here $\sigma_+ = |1\rangle\langle 0|$ is the raising operator and $\sigma_- = \sigma_+^*$). Then $\|\mathcal{M}(\rho) - \mathcal{M}(\sigma)\|_p = 2^{1/p}$ is larger than $\|\rho - \sigma\|_p = 2^{2/p-1}$.

Proposition 7.1.2. [154] *The trace distance d_1 is contractive.*

Proof. : Let $R = \rho - \sigma = R_+ - R_-$, $R_\pm = (|R| \pm R)/2 \geq 0$ being the positive and negative parts of R. Then $\|R\|_1 = \operatorname{tr}(R_+ + R_-) = 2\operatorname{tr}(R_+)$ because $\operatorname{tr}(R) = \operatorname{tr}(R_+) - \operatorname{tr}(R_-) = 0$. Since $\mathcal{M}$ is trace preserving and CP, one has similarly $\|\mathcal{M}(R)\|_1 = 2\operatorname{tr}(\mathcal{M}(R)_+)$ and $\mathcal{M}(R) = \mathcal{M}(R_+) - \mathcal{M}(R_-) \leq \mathcal{M}(R_+)$. Using $\mathcal{M}(R)_+ = P_+\mathcal{M}(R)P_+$ with P_+ the spectral projector of $\mathcal{M}(R)$ on $[0, \infty)$, one gets

$$\|\mathcal{M}(R)\|_1 = 2\operatorname{tr}[P_+\mathcal{M}(R)P_+] \leq 2\operatorname{tr}[P_+\mathcal{M}(R_+)P_+] \leq 2\operatorname{tr}[\mathcal{M}(R_+)] = 2\operatorname{tr}[R_+] = \|R\|_1 \ , \tag{172}$$

hence the result. □

A distance d on $\mathcal{E}(\mathcal{H})$ is *jointly convex* if for any state ensembles $\{\rho_i, p_i\}$ and $\{\sigma_i, p_i\}$ with the same probabilities p_i,

$$d\Big(\sum_i p_i\rho_i, \sum_i p_i\sigma_i\Big) \leq \sum_i p_i d(\rho_i, \sigma_i) \ . \tag{173}$$

Since they are associated to a norm, the distances d_p are jointly convex for any $p \geq 1$.

7.2 The Bures distance

We now introduce the Bures distance $d_{\rm B}$. This distance is contractive like d_1. It was first considered by Bures in the context of infinite products of von Neumann algebras [35] (see also [10]) and was later studied in a series of papers by Uhlmann [178, 180, 182]. Uhlmann used it to define parallel transport and related it to the fidelity generalizing the usual fidelity $|\langle\psi|\phi\rangle|^2$ between pure states. Indeed, $d_{\rm B}$ is an extension to mixed states of the Fubini-Study distance on the projective space $P\mathcal{H}$ of pure states,

$$d_{\rm FS}(\rho_\psi, \sigma_\phi) = \inf_{\|\psi_\theta\| = \|\phi_\delta\| = 1} \big\| |\psi_\theta\rangle - |\phi_\delta\rangle \big\| = \big(2 - 2|\langle\psi|\phi\rangle|\big)^{\frac{1}{2}} \ , \tag{174}$$

where the infimum in the second member is over all representatives $|\psi_\theta\rangle$ of $\rho_\psi \in P\mathcal{H}$ and $|\phi_\delta\rangle$ of $\sigma_\phi \in P\mathcal{H}$ (i.e., $|\psi_\theta\rangle = e^{i\theta}|\psi\rangle$ and $|\phi_\delta\rangle = e^{i\delta}|\phi\rangle$). Observe that the third member is independent of these representatives. For two mixed states ρ and σ in $\mathcal{E}(\mathcal{H})$, one can define analogously [180, 99]

$$d_{\rm B}(\rho, \sigma) = \inf_{A,B} d_2(A, B) \ , \tag{175}$$

where the infimum is over all Hilbert-Schmidt matrices $A \in \mathcal{B}(\mathcal{H})$ and $B \in \mathcal{B}(\mathcal{H})$ satisfying $AA^* = \rho$ and $BB^* = \sigma$. Such matrices are given by $A = \sqrt{\rho}V$ and $B = \sqrt{\sigma}W$ for some unitaries V and W (polar decompositions). If $\rho = \rho_\psi$ and $\sigma = \sigma_\phi$ are pure states, then $A = |\psi\rangle\langle\mu|$ and $B = |\phi\rangle\langle\nu|$ with $\|\mu\| = \|\nu\| = 1$, so that (175) reduces to the Fubini-Study distance (174).

Let us show that $d_{\rm B}$ defines a distance on the set of quantum states $\mathcal{E}(\mathcal{H})$. It is immediate on (175) that $d_{\rm B}$ is non-negative and symmetric, and that $d_{\rm B}(\rho, \sigma) = 0$ if and only if $\rho = \sigma$. To prove that $d_{\rm B}$ satisfies the triangle inequality, let us first observe that by the polar decomposition and the invariance property $d_2(AV, BV) = d_2(A, B)$ of the Hilbert-Schmidt distance (here V is any unitary operator), the definition (175) can be rewritten as $d_{\rm B}(\rho, \sigma) = \inf_U d_2(\sqrt{\rho}, \sqrt{\sigma}U)$ with an infimum over all unitaries U. Let ρ, σ, and τ be three states of $\mathcal{E}(\mathcal{H})$. The triangle inequality for d_2 and the aforementioned invariance property yield

$$\begin{aligned} d_{\rm B}(\rho, \tau) &\leq \inf_{U,V} \big\{ d_2(\sqrt{\rho}, \sqrt{\sigma}U) + d_2(\sqrt{\sigma}U, \sqrt{\tau}V) \big\} = \inf_U d_2(\sqrt{\rho}, \sqrt{\sigma}U) + \inf_W d_2(\sqrt{\sigma}, \sqrt{\tau}W) \\ &= d_{\rm B}(\rho, \sigma) + d_{\rm B}(\sigma, \tau) \ . \end{aligned} \tag{176}$$

Hence the map $(\rho, \sigma) \mapsto d_{\rm B}(\rho, \sigma)$ defines a distance on the set of quantum states.

For any states ρ and $\sigma \in \mathcal{E}(\mathcal{H})$, the right-hand side of (175) is given by

$$\big(2 - 2\sup_U \mathrm{Re}\ \mathrm{tr}(U\sqrt{\rho}\sqrt{\sigma})\big)^{\frac{1}{2}} \tag{177}$$

with a supremum over all unitaries $U = WV^*$. This supremum is equal to $\|\sqrt{\rho}\sqrt{\sigma}\|_1$ and is attained if and only if $UU_0|\sqrt{\rho}\sqrt{\sigma}|^{\frac{1}{2}} = |\sqrt{\rho}\sqrt{\sigma}|^{\frac{1}{2}}$, where U_0 is such that $\sqrt{\rho}\sqrt{\sigma} = U_0|\sqrt{\rho}\sqrt{\sigma}|$ (see Sec. 2.1). Equivalently, the infimum in (175) is attained if and only if the parallel transport condition $A^*B \geq 0$ holds. We obtain the two following equivalent definitions of the Bures distance.

Definition 7.2.1. *The Bures distance on the set of quantum states $\mathcal{E}(\mathcal{H})$ is the distance d_{B} defined by (175). Equivalently, for any states $\rho, \sigma \in \mathcal{E}(\mathcal{H})$,*

$$d_{\mathrm{B}}(\rho,\sigma) = \big(2 - 2\sqrt{F(\rho,\sigma)}\big)^{\frac{1}{2}} \tag{178}$$

where the Uhlmann fidelity is defined by

$$F(\rho,\sigma) = \|\sqrt{\rho}\sqrt{\sigma}\|_1^2 = \Big(\mathrm{tr}\big[(\sqrt{\sigma}\rho\sqrt{\sigma})^{\frac{1}{2}}\big]\Big)^2 . \tag{179}$$

Thus d_{B} takes values in $[0, \sqrt{2}]$.

The fidelity $F(\rho,\sigma)$ is symmetric in (ρ,σ) and belongs to the interval $[0,1]$. It is clearly a generalization of the usual pure state fidelity $F(|\psi\rangle, |\phi\rangle) = |\langle\psi|\phi\rangle|^2$. If σ_ϕ is pure, then

$$F(\rho,\sigma_\phi) = \langle\phi|\rho|\phi\rangle \tag{180}$$

for any $\rho \in \mathcal{E}(\mathcal{H})$.

The following theorem will prove to be quite useful for establishing various properties of the Bures distance.

Theorem 7.2.2. (Uhlmann [178]) *Let ρ, $\sigma \in \mathcal{E}(\mathcal{H})$ and $|\Psi\rangle$ be a purification of ρ on the space $\mathcal{H}\otimes\mathcal{K}$, with* $\dim\mathcal{K} \geq \dim\mathcal{H}$. *Then*

$$F(\rho,\sigma) = \max_{|\Phi\rangle} |\langle\Psi|\Phi\rangle|^2 \tag{181}$$

where the maximum is over all purifications $|\Phi\rangle$ of σ on $\mathcal{H}\otimes\mathcal{K}$.

Proof. Let us first assume $\mathcal{K} = \mathcal{H}$. Then (181) follows from the definition (175) of the Bures distance and the fact that the map $A \mapsto |\Psi_A\rangle = \sum_{i,j}\langle i|A|j\rangle|i\rangle|j\rangle$ is an isometry between $\mathcal{B}(\mathcal{H})$ (endowed with the Hilbert-Schmidt norm $\|\cdot\|_2$) and $\mathcal{H}\otimes\mathcal{H}$ (here $\{|i\rangle\}$ is some fixed orthonormal basis of $\mathcal{H}$, see Sec. 2.1). Indeed, one easily checks that $\rho = AA^*$ if and only if $|\Psi_A\rangle$ is a purification of ρ on $\mathcal{H}\otimes\mathcal{H}$. Hence, using (174), (175), and the invariance property of d_2 with respect to right multiplications by unitaries, one has

$$d_{\mathrm{B}}(\rho,\sigma)^2 = \inf_B \big\|A-B\big\|_2^2 = \inf_{|\Phi_B\rangle} \big\||\Psi_A\rangle - |\Phi_B\rangle\big\|^2 = \inf_{|\Phi_B\rangle} d_{\mathrm{FS}}\big(|\Psi_A\rangle, |\Phi_B\rangle\big)^2 = 2 - 2\sup_{|\Phi_B\rangle} |\langle\Psi_A|\Phi_B\rangle| \ , \tag{182}$$

where the infimum and supremum are over all purifications $|\Phi_B\rangle$ of σ on $\mathcal{H}\otimes\mathcal{H}$, and are actually minimum and maximum.

If $\dim\mathcal{K} > \dim\mathcal{H}$, we extend ρ and σ to a larger space $\mathcal{H}' \simeq \mathcal{K}$ by adding to them new orthonormal eigenvectors with zero eigenvalues. As is clear from (178)-(179), this does not change the distance, hence $d_{\mathrm{B}}(\rho,\sigma) = \inf_{A',B'} d_2(A',B')$ with an infimum over all $A',B' \in \mathcal{B}(\mathcal{H}')$ such that $A'(A')^*$ and $B'(B')^*$ are equal to the aforementioned extensions of ρ and σ. But A' and B' can be viewed as operators from $\mathcal{K}$ to $\mathcal{H}$ since they have ranges $\operatorname{ran} A' = \ker(A')^*$ and $\operatorname{ran} B' = \ker(B')^*$ included in $\mathcal{H}$. Thus, one can take the infimum in (175) over all $A,B \in \mathcal{B}(\mathcal{K},\mathcal{H})$ such that $AA^* = \rho$ and $BB^* = \sigma$, without changing the result. The formula (181) then follows from the same argument as above, using the fact that $A \mapsto |\Psi_A\rangle$ is an isometry from $\mathcal{B}(\mathcal{K},\mathcal{H})$ to $\mathcal{H} \otimes \mathcal{K}$ (Sec. 2.1). $\square$

A direct proof of (181) from the definition (179) of the fidelity has been given Ref. [106] (see also [134]).

Corollary 7.2.3. *The distance* d_{B} *on the sets of quantum states is contractive. Moreover,* d_{B}^2 *is jointly convex.*

Note that d_{B} is not jointly convex. One gets a counter-example by choosing $\rho_0 = \sigma_0 = |0\rangle\langle 0|$, $\rho_1 = |1\rangle\langle 1|$, $\sigma_1 = |2\rangle\langle 2|$, and $p_0 = p_1 = 1/2$, $\{|0\rangle, |1\rangle, |2\rangle\}$ being an orthonormal family in $\mathcal{H}$.

It is clear on (179) that $F(\rho,\sigma) = 0$ if and only if ρ and σ have orthogonal supports, $\operatorname{ran}\rho \perp \operatorname{ran}\sigma$. Therefore, two states ρ and σ have a maximal distance $d_{\mathrm{B}}(\rho,\sigma) = \sqrt{2}$ if they are orthogonal and thus perfectly distinguishable.

Proof. To show the contractivity, it is enough to check that for any quantum operation $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ and any states $\rho, \sigma \in \mathcal{E}(\mathcal{H})$,

$$F(\mathcal{M}(\rho), \mathcal{M}(\sigma)) \geq F(\rho,\sigma) \; . \tag{183}$$

This property of the fidelity is a consequence of the contractivity of the relative Rényi entropy for $\alpha = 1/2$ (Theorem 6.3.1*(v)*). It is, however, instructive to re-derive this result from Theorem 7.2.2. According to this theorem, there exist some purifications $|\Psi\rangle$ and $|\Phi\rangle$ of ρ and σ on $\mathcal{H} \otimes \mathcal{K}$ such that $F(\rho,\sigma) = |\langle\Psi|\Phi\rangle|^2$. Now, thanks to (34) one obtains some purifications $|\Psi_{\mathcal{M}}\rangle = 1_{\mathcal{K}} \otimes U|\Psi\rangle|\epsilon_0\rangle$ of $\mathcal{M}(\rho)$ and $|\Phi_{\mathcal{M}}\rangle = 1_{\mathcal{K}} \otimes U|\Phi\rangle|\epsilon_0\rangle$ of $\mathcal{M}(\sigma)$ on $\mathcal{K} \otimes \mathcal{H}' \otimes \mathcal{H}'_{\mathsf{E}}$, with $|\epsilon_0\rangle \in \mathcal{H}_{\mathsf{E}}$ and $U : \mathcal{H} \otimes \mathcal{H}_{\mathsf{E}} \to \mathcal{H}' \otimes \mathcal{H}'_{\mathsf{E}}$ unitary. Thus

$$F(\mathcal{M}(\rho), \mathcal{M}(\sigma)) \geq |\langle\Psi_{\mathcal{M}}|\Phi_{\mathcal{M}}\rangle|^2 = |\langle\Psi|\Phi\rangle|^2 = F(\rho,\sigma) \; . \tag{184}$$

The joint convexity of d_{B}^2 is a consequence of the bound[21]

$$\sqrt{F\Big(\sum_i p_i\rho_i, \sum_i q_i\sigma_i\Big)} \geq \sum_i \sqrt{p_i q_i}\sqrt{F(\rho_i,\sigma_i)} \; , \tag{185}$$

where $\{\rho_i, p_i\}$ and $\{\sigma_i, q_i\}$ are arbitrary ensembles in $\mathcal{E}(\mathcal{H})$. Note that the statement (185) is slightly more general than the joint concavity of $\sqrt{F(\rho,\sigma)}$ proven in Sec. 6.3 (Theorem 6.3.1*(iv)*). To show that (185) is true, we introduce as before some purifications $|\Psi_i\rangle$ of ρ_i and $|\Phi_i\rangle$ of σ_i on $\mathcal{H} \otimes \mathcal{H}$ such that $\sqrt{F(\rho_i,\sigma_i)} = \langle\Psi_i|\Phi_i\rangle$. Let us define the vectors

$$|\Psi\rangle = \sum_i \sqrt{p_i}|\Psi_i\rangle|\epsilon_i\rangle \quad , \quad |\Phi\rangle = \sum_i \sqrt{q_i}|\Phi_i\rangle|\epsilon_i\rangle \tag{186}$$

[21] Note that one cannot replace $\sqrt{F}$ by F in this inequality, that is, $F(\rho,\sigma)$ is not jointly concave (one can take the same counter-example as that given above for d_{B}). However, by a slight modification of the proof of Corollary 7.2.3 one can show that $\rho \mapsto F(\rho,\sigma)$ and $\sigma \mapsto F(\rho,\sigma)$ are concave. In their book [134], Nielsen and Chuang define the fidelity as the square root of (179). This must be kept in mind when comparing the results in this monograph with those of this article.

in $\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}_{\mathsf{E}}$, where $\mathcal{H}_{\mathsf{E}}$ is an auxiliary Hilbert space and $\{|\epsilon_i\rangle\}$ is an orthonormal basis of $\mathcal{H}_{\mathsf{E}}$. Then $|\Psi\rangle$ and $|\Phi\rangle$ are purifications of $\rho = \sum_i p_i \rho_i$ and $\sigma = \sum_i q_i \sigma_i$, respectively. One infers from Theorem 7.2.2 that

$$\sqrt{F(\rho,\sigma)} \geq |\langle\Psi|\Phi\rangle| = \sum_i \sqrt{p_i q_i} \langle\Psi_i|\Phi_i\rangle = \sum_i \sqrt{p_i q_i} \sqrt{F(\rho_i,\sigma_i)} \,. \tag{187}$$

This complete the proof of the corollary. □

Remark 7.2.4. *A consequence of (47) and (180) and of the monotonicity of the fidelity F with respect to partial trace operations (see (183)) is that the entanglement fidelity $F_{\mathrm{e}}(\rho, \mathcal{M})$ of a state ρ with respect to a quantum operation $\mathcal{M}$ satisfies*

$$F_{\mathrm{e}}(\rho, \mathcal{M}) \leq F(\rho, \mathcal{M}(\rho)) \,. \tag{188}$$

Remark 7.2.5. *As the fidelity satisfies $F(\rho\otimes\rho', \sigma\otimes\sigma') = F(\rho,\sigma)F(\rho',\sigma')$, the Bures distance increases by taking tensor products, $d_{\mathrm{B}}(\rho\otimes\rho', \sigma\otimes\sigma') \geq d_{\mathrm{B}}(\rho,\sigma)$ for any $\rho, \sigma \in \mathcal{E}(\mathcal{H})$, $\rho', \sigma' \in \mathcal{E}(\mathcal{H}')$, with equality if and only if $\rho' = \sigma'$. This has to be contrasted with the trace distance, which does not enjoy this property.*

In the two following subsections we collect some important properties of the Bures distance. We refer the reader to the monographs [22, 134] for a list of names to which these properties should be attached.

7.3 Bures distance and statistical distance in classical probability

The restriction of a distance d on $\mathcal{E}(\mathcal{H})$ to all density matrices commuting with a given state ρ_0 defines a distance on the simplex $\mathcal{E}_{\mathrm{clas}} = \{\mathbf{p} \in \mathbb{R}_+^n; \sum_i p_i = 1\}$ of classical probabilities on the finite space $\{1, 2, \ldots, n\}$. In particular, if ρ and σ are two commuting states with spectral decompositions $\rho = \sum_k p_k |k\rangle\langle k|$ and $\sigma = \sum_k q_k |k\rangle\langle k|$, then

$$d_1(\rho,\sigma) = d_1^{\mathrm{clas}}(\mathbf{p},\mathbf{q}) = \sum_{k=1}^{n} |p_k - q_k|$$

is the ℓ^1-distance, and

$$d_{\mathrm{B}}(\rho,\sigma) = d_{\mathrm{H}}^{\mathrm{clas}}(\mathbf{p},\mathbf{q}) = \left(\sum_{k=1}^{n} (\sqrt{p_k} - \sqrt{q_k})^2\right)^{\frac{1}{2}} = \Big(2 - 2\sum_{k=1}^{n} \sqrt{p_k q_k}\Big)^{\frac{1}{2}} \tag{189}$$

is the Hellinger distance. A distance closely related to $d_{\mathrm{H}}^{\mathrm{clas}}$ is the so-called statistical distance $\Theta^{\mathrm{clas}}(\mathbf{p},\mathbf{q}) = \arccos(1 - d_{\mathrm{H}}^{\mathrm{clas}}(\mathbf{p},\mathbf{q})^2/2)$, i.e., the angle between the vectors $\mathbf{x} = (\sqrt{p_k})_{k=1}^n$ and $\mathbf{y} = (\sqrt{q_k})_{k=1}^n$ on the unit sphere. Given two non-commuting states ρ and σ, one can consider the distance $d^{\mathrm{clas}}(\mathbf{p},\mathbf{q})$ between the outcome probabilities $\mathbf{p}$ and $\mathbf{q}$ of a measurement performed on the system in states ρ and σ, respectively. It is natural to ask whether there is a relation between $d(\rho,\sigma)$ and the supremum of $d^{\mathrm{clas}}(\mathbf{p},\mathbf{q})$ over all measurements.

Proposition 7.3.1. *For any $\rho, \sigma \in \mathcal{E}(\mathcal{H})$,*

$$d_1(\rho,\sigma) = \sup_{\{M_i\}} d_1^{\rm clas}(\mathbf{p},\mathbf{q}) \quad , \quad d_{\rm B}(\rho,\sigma) = \sup_{\{M_i\}} d_{\rm H}^{\rm clas}(\mathbf{p},\mathbf{q}) \; , \tag{190}$$

where the suprema are over all POVMs $\{M_i\}$ and $p_i = \operatorname{tr}(M_i\rho)$ (respectively $q_i = \operatorname{tr}(M_i\sigma)$) is the probability of the measurement outcome i in the state ρ (respectively σ). Moreover, the suprema are achieved for von Neumann measurements with rank-one projectors $M_i = |i\rangle\langle i|$.

Remark 7.3.2. *It is worth pointing out that Proposition 7.3.1 provides an alternative proof of the contractivity of d_1 and $d_{\rm B}$ under quantum operations, which runs as follows. First note that any quantum operation $\mathcal{M}$ followed by a measurement with POVM $\{M_i\}$ is equivalent to a single measurement with POVM $\{\mathcal{M}^*(M_i)\}$, the probability p_i of outcome i being the same in the two cases since $p_i = \operatorname{tr}[M_i\mathcal{M}(\rho)] = \operatorname{tr}[\mathcal{M}^*(M_i)\rho]$ by definition of the dual map. Thus, if $\{M_i^{\rm opt}\}$ is an optimal POVM achieving the supremum of the classical distance and $\mathbf{p}^{\rm opt}$ and $\mathbf{q}^{\rm opt}$ are the corresponding probabilities for the states $\mathcal{M}(\rho)$ and $\mathcal{M}(\sigma)$, we deduce from Proposition 7.3.1 that*

$$d_{\rm B}(\mathcal{M}(\rho),\mathcal{M}(\sigma)) = d_{\rm H}^{\rm clas}(\mathbf{p}^{\rm opt},\mathbf{q}^{\rm opt}) = d_{\rm H}^{\rm clas}\big(\big\{ \operatorname{tr}[\mathcal{M}^*(M_i^{\rm opt})\rho]\big\},\big\{ \operatorname{tr}[\mathcal{M}^*(M_i^{\rm opt})\sigma]\big\}\big) \leq d_{\rm B}(\rho,\sigma) \; , \tag{191}$$

with a similar inequality for d_1.

Proof. We leave the justification of the first identity to the reader. It can be obtained by following similar arguments as in the proof of Proposition 7.1.2 (see [134]). Let us show the second identity. Given a POVM $\{M_i\}$, by taking advantage of the definition (179) of the fidelity, the polar decomposition $\sqrt{\rho}\sqrt{\sigma} = U|\sqrt{\rho}\sqrt{\sigma}|$, and the identity $\sum_i M_i = 1$, one gets

$$\sqrt{F(\rho,\sigma)} = \sum_i \operatorname{tr}(U^*\sqrt{\rho}\sqrt{M_i}\sqrt{M_i}\sqrt{\sigma}) \;\leq\; \sum_i \sqrt{p_i q_i} \; . \tag{192}$$

The upper bound comes from the Cauchy-Schwarz inequality. It remains to show that this bound can be attained for an appropriate choice of POVM. The Cauchy-Schwarz inequality holds with equality if and only if $\sqrt{M_i}\sqrt{\rho}U = \lambda_i\sqrt{M_i}\sqrt{\sigma}$ with $\lambda_i \geq 0$. Assuming $\sigma > 0$ and observing that $\sqrt{\sigma}\sqrt{\rho} = (\sqrt{\rho}\sqrt{\sigma})^* = |\sqrt{\rho}\sqrt{\sigma}|U^*$ and thus $\sqrt{\rho}U = \sigma^{-\frac{1}{2}}|\sqrt{\rho}\sqrt{\sigma}|$, this identity can be recast as

$$\sqrt{M_i}(R-\lambda_i) = 0 \quad \text{ with } \quad R = \sigma^{-\frac{1}{2}}|\sqrt{\rho}\sqrt{\sigma}|\sigma^{-\frac{1}{2}} \; . \tag{193}$$

Let $R = \sum_i r_i |i\rangle\langle i|$ be a spectral decomposition of the self-adjoint matrix R. Taking M_i to be the von Neumann projector $M_i = |i\rangle\langle i|$ and $\lambda_i = r_i$, we find that (193) is satisfied for all i. Thus $\sqrt{F(\rho,\sigma)}$ is equal to the right-hand side of (192). If σ is not invertible it can be approached by invertible density matrices $\sigma_\varepsilon = (1-\varepsilon)\sigma + \varepsilon$, $\varepsilon > 0$, and the result follows by continuity. □

The proof shows that the best measurement to discriminate the two states ρ and σ in a statistical sense, that is, the measurement which maximizes the Hellinger distance between the distributions of outcomes $\mathbf{p}$ and $\mathbf{q}$ in these two states, is the von Neumann measurement of the observable R defined in (193) [67]. If ρ and σ are invertible, this optimal measurement is unique up to rotations in the corresponding eigenspace(s) if R has some degenerate eigenvalue(s). Although the optimal measurement has to be symmetric under the exchange of ρ and σ, this is not obvious at first glance from the expression of R. Let us show that the exchange of ρ and σ in (193) amounts to invert R. Actually, one deduces from the polar decomposition $\sqrt{\rho}\sqrt{\sigma} = U|\sqrt{\rho}\sqrt{\sigma}|$ that $|\sqrt{\sigma}\sqrt{\rho}| = U|\sqrt{\rho}\sqrt{\sigma}|U^*$

(recall that if $A = U|A|$ with $|A| = \sqrt{A^* A}$ then $|A^*| = U|A|U^*$). Similarly, one has $\rho^{\frac{1}{2}} U = \sigma^{-\frac{1}{2}} |\sqrt{\rho}\sqrt{\sigma}|$ (see above). Thus, if ρ and σ are invertible then

$$\begin{aligned} \rho^{-\frac{1}{2}} |\sqrt{\sigma}\sqrt{\rho}| \rho^{-\frac{1}{2}} &= \rho^{-\frac{1}{2}} U |\sqrt{\rho}\sqrt{\sigma}| U^* \rho^{-\frac{1}{2}} = \left[\rho^{\frac{1}{2}} U |\sqrt{\rho}\sqrt{\sigma}|^{-1} U^* \rho^{\frac{1}{2}}\right]^{-1} \\ &= \left[\sigma^{-\frac{1}{2}} |\sqrt{\rho}\sqrt{\sigma}|\, |\sqrt{\rho}\sqrt{\sigma}|^{-1} |\sqrt{\rho}\sqrt{\sigma}| \sigma^{-\frac{1}{2}}\right]^{-1} = R^{-1} \,. \end{aligned} \tag{194}$$

Since R and R^{-1} have same eigenbasis, one gets the same optimal measurement $\{M_i^{\rm opt} = |i\rangle\langle i|\}$ by exchanging ρ and σ. It is also worth mentioning that the equality

$$\sqrt{M_i^{\rm opt}} \rho \sqrt{M_i^{\rm opt}} = r_i^2 \sqrt{M_i^{\rm opt}} \sigma \sqrt{M_i^{\rm opt}} \,, \tag{195}$$

which follows from the optimality condition $\sqrt{M_i^{\rm opt}} \sqrt{\rho}\, U = r_i \sqrt{M_i^{\rm opt}} \sqrt{\sigma}$ in the proof of Proposition 7.3.1, implies that the optimal outcome distributions are related by

$$p_i^{\rm opt} = r_i^2 q_i^{\rm opt} \tag{196}$$

with r_i the eigenvalues of R.

Remark 7.3.3. *A nice application of the second equality in (190) is a strong version of the no-cloning theorem for mixed states called the no-broadcasting theorem [20]. This theorem states that, given two non-commutating states ρ_0 and ρ_1 of a system S and a register R prepared initially in state τ, there is no physical evolution described by a quantum operation $\mathcal{M}$ on SR such that the final reduced states of S and R are equal to ρ_s for $s \in \{0,1\}$, that is, ${\rm tr}_{\mathsf{R}}(\omega_s') = {\rm tr}_{\mathsf{S}}(\omega_s') = \rho_s$, where $\omega_s = \rho_s \otimes \tau$ and $\omega_s' = \mathcal{M}(\omega_s)$ are the initial and final states, respectively. Note that one does not assume that ω_s' is a product state $\rho_s \otimes \rho_s$ as in the no-cloning theorem. The proof of this result is based on the following observation. Let $\{M_i^{\rm opt}\}$ be an optimal POVM achieving the supremum in (190). One deduces from*

$$p_{i|s}^{\rm opt} = {\rm tr}_{\mathsf{S}} \left[M_i^{\rm opt} \rho_s\right] = {\rm tr}_{\mathsf{SR}} \left[M_i^{\rm opt} \otimes 1\, \omega_s'\right] = {\rm tr}_{\mathsf{SR}} \left[1 \otimes M_i^{\rm opt}\, \omega_s'\right] \quad , \; s = 0, 1 \tag{197}$$

that

$$\begin{aligned} d_{\rm B}(\rho_0, \rho_1) = d_{\rm H}^{\rm clas}(\{p_{i|0}^{\rm opt}\}, \{p_{i|1}^{\rm opt}\}) &\leq \sup_{POVM\,\{P_i\}} d_{\rm H}^{\rm clas}(\{{\rm tr}_{\mathsf{SR}}[P_i\, \omega_0']\}, \{{\rm tr}_{\mathsf{SR}}[P_i\, \omega_1']\}) \\ &= d_{\rm B}(\mathcal{M}(\rho_0), \mathcal{M}(\rho_1)) \,. \end{aligned} \tag{198}$$

The contractivity of $d_{\rm B}$ imposes the reverse inequality, hence $d_{\rm B}(\rho_0, \rho_1) = d_{\rm B}(\mathcal{M}(\rho_0), \mathcal{M}(\rho_1))$ and the supremum in (198) is achieved for the POVMs $\{M_i^{\rm opt} \otimes 1\}$ and $\{1 \otimes M_i^{\rm opt}\}$. Using the necessary and sufficient conditions in the form (193) insuring that these POVMs are optimal, one can show that this implies that ρ_0 and ρ_1 commute (see [20] for more detail).

7.4 Comparison of the Bures, trace, and quantum Hellinger distances

Much as for the quantum relative Rényi entropies (Sec. 6.3), one may define another distance on $\mathcal{E}(\mathcal{H})$ which also reduces to the Hellinger distance $d_{\rm H}^{\rm clas}$ for commuting matrices, by setting

$$d_{\rm H}(\rho, \sigma) = d_2(\sqrt{\rho}, \sqrt{\sigma}) = \left(2 - 2\sqrt{F_{\frac{1}{2}}^{(\rm n)}(\rho||\sigma)}\right)^{\frac{1}{2}} \,, \tag{199}$$

where $F_\alpha^{(\rm n)}(\rho||\sigma)$ is the fidelity associated to the normal-ordered α-entropy (143), namely,

$$F_\alpha^{(\rm n)}(\rho||\sigma) = \left({\rm tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]\right)^{\frac{1}{\alpha}} = e^{-\beta S_\alpha^{(\rm n)}(\rho||\sigma)} \quad , \quad \beta = \frac{1-\alpha}{\alpha} \,. \tag{200}$$

This distance is sometimes called the quantum Hellinger distance. Thanks to Lieb's concavity theorem (Lemma 6.3.2), $F_\alpha^{(\mathrm{n})}(\rho||\sigma)^\alpha$ is jointly concave in (ρ,σ) for all $\alpha \in (0,1)$. Consequently, the square Hellinger distance $d_{\mathrm{H}}(\rho,\sigma)^2$ is jointly convex, just as $d_{\mathrm{B}}(\rho,\sigma)^2$. From Proposition 6.2.2 one then deduces that d_{H} is contractive. It is worth noting that d_{H} does not coincide with the Fubini-Study distance (174) for pure states (in fact, one finds $F_{1/2}^{(\mathrm{n})}(\rho_\psi||\sigma_\phi) = |\langle\psi|\phi\rangle|^4$).

The next result shows that the Bures, trace, and quantum Hellinger distances are equivalent and gives explicit bounds of one distance in terms of the others.

Proposition 7.4.1. *For any $\rho,\sigma \in \mathcal{E}(\mathcal{H})$, one has*

$$d_{\mathrm{B}}(\rho,\sigma) \leq d_{\mathrm{H}}(\rho,\sigma) \leq \sqrt{2}\, d_{\mathrm{B}}(\rho,\sigma) \tag{201}$$

$$d_{\mathrm{H}}(\rho,\sigma)^2 \leq d_1(\rho,\sigma) \leq 2\Big\{1 - \Big(1 - \frac{1}{2} d_{\mathrm{B}}(\rho,\sigma)^2\Big)^2\Big\}^{\frac{1}{2}} . \tag{202}$$

The last inequality in (202) is saturated for pure states.

The bounds $d_{\mathrm{B}}(\rho,\sigma)^2 \leq d_1(\rho,\sigma)$ and $d_{\mathrm{H}}(\rho,\sigma)^2 \leq d_1(\rho,\sigma)$, which are consequences of (201) and (202), have been first proven in the C^*-algebra setting by Araki [10] and Holevo [88], respectively. An upper bound on $d_1(\rho,\sigma)$ similar to the one in (202) but with d_{B} replaced by d_{H} (which is weaker than the bound in (202) because of (201)) has been also derived by Holevo [88]. Note that the last bound in (202) implies that $d_1(\rho,\sigma) \leq 2 d_{\mathrm{B}}(\rho,\sigma)$.

Proof. For any $\rho,\sigma \in \mathcal{E}(\mathcal{H})$, one finds by comparing (175) and (199) that $d_{\mathrm{B}}(\rho,\sigma) \leq d_{\mathrm{H}}(\rho,\sigma)$. The second bound in (201) is a consequence of the inequalities (224) below on the Bures and Hellinger metrics. The first bound in (202) can be obtained by using the identity (B6) from Appendix B with $s = 1/2$, $A = \rho$, and $B = \sigma$. To prove the last bound in (202), we first argue that if $\rho_\psi = |\psi\rangle\langle\psi|$ and $\sigma_\phi = |\phi\rangle\langle\phi|$ are pure states, then $d_1(\rho_\psi,\sigma_\phi) = 2\sqrt{1 - F(\rho_\psi,\sigma_\phi)}$, showing that this bound holds with equality. Actually, let $|\phi\rangle = \cos\theta|\psi\rangle + e^{\mathrm{i}\delta}\sin\theta|\psi^\perp\rangle$, where $\theta,\delta \in [0,2\pi)$ and $|\psi^\perp\rangle$ is a unit vector orthogonal to $|\psi\rangle$. Since $\rho_\psi - \sigma_\phi$ has non-vanishing eigenvalues $\pm\sin\theta$, one has $d_1(\rho_\psi,\sigma_\phi) = 2|\sin\theta|$. But $F(\rho_\psi,\sigma_\phi) = \cos^2\theta$, hence the aforementioned statement is true. It then follows from Theorem 7.2.2 and from the contractivity of the trace distance with respect to partial trace operations (Proposition 7.1.2) that for arbitrary ρ and $\sigma \in \mathcal{E}(\mathcal{H})$,

$$d_1(\rho,\sigma) \leq 2\sqrt{1 - F(\rho,\sigma)} . \tag{203}$$

This concludes the proof. □

The following bound on the relative entropy can be obtained from (150), (146), and Proposition 6.3.4

$$S(\rho||\sigma) \geq -2\ln\Big(1 - \frac{1}{2} d_{\mathrm{B}}(\rho,\sigma)^2\Big) \geq -\ln\Big(1 - \frac{1}{4} d_1(\rho,\sigma)^2\Big) . \tag{204}$$

Remark 7.4.2. *By taking advantage of the inequality $F(\rho,\sigma) \geq \mathrm{tr}(\rho\sigma)$, which follows from (179) and the norm inequality $\|A\|_1 \geq \|A\|_2$, one can establish another bound on $S(\rho||\sigma)$ in terms of the fidelity, which reads [173]*

$$S(\rho||\sigma) \geq -S(\rho) - \ln F(\rho,\sigma) . \tag{205}$$

Remark 7.4.3. *The bound $d_{\mathrm{B}}(\rho,\sigma)^2 \leq d_1(\rho,\sigma)$ can be derived directly (without relying on the similar inequality for the quantum Hellinger distance) by using Proposition 7.3.1 [134].*

Remark 7.4.4. *The formula*

$$F(\rho,\sigma) = \frac{1}{4}\inf_{H>0}\left\{ \operatorname{tr}(H\rho) + \operatorname{tr}(H^{-1}\sigma)\right\}^2 = \inf_{H>0}\left\{\operatorname{tr}(H\rho)\operatorname{tr}(H^{-1}\sigma)\right\} \tag{206}$$

can be easily proven with the help of Lemma 6.3.3 and Theorem 7.2.2. The last expression is due to Alberti [6].

Remark 7.4.5. *It may be useful to derive bounds on the fidelity involving traces of polynomials in ρ and σ, instead of traces of polynomials in $\sqrt{\rho}$ and $\sqrt{\sigma}$ or of the trace distance, which are harder to compute. An example of such lower and upper bounds is [125]:*

$$\operatorname{tr}[\rho\sigma] + \sqrt{2(\operatorname{tr}[\rho\sigma])^2 - 2\operatorname{tr}[(\rho\sigma)^2]} \leq F(\rho,\sigma) \leq \operatorname{tr}[\rho\sigma] + \sqrt{(1-\operatorname{tr}[\rho^2])(1-\operatorname{tr}[\sigma^2])}\ . \tag{207}$$

Remark 7.4.6. *We are now in position to show without much effort several results of Sec. 5.2.*

(a) The upper bound (77) on the optimal success probability $P_{\rm S,u}^{\rm opt}$ in unambiguous discrimination of two mixed states can be established from Uhlmann's theorem, formula (70), and the fact that $P_{\rm S,u}^{\rm opt}(\{\rho_i,\eta_i\}) \leq P_{\rm S,u}^{\rm opt}(\{|\Psi_i\rangle,\eta_i\})$, where $|\Psi_i\rangle$ is a purification of ρ_i for any i [153].

(b) It is instructive to derive in the special case of $m=2$ states the lower bound on $P_{\rm S,a}^{\rm opt}$ given in Proposition 5.5.1 by using the Helstrom formula (68), the fact that $\operatorname{tr}(|\Lambda|) \geq \sum_i |\langle i|\Lambda|i\rangle|$ for any orthonormal basis $\{|i\rangle\}$, and Proposition 7.3.1 [28].

(c) The Uhlmann theorem gives an efficient way to calculate the fidelity between the two states (78) (the result is $F(\rho_{\rm eq},\rho_{\rm diff}) = |\langle\psi_1|\psi_2\rangle|^2$).

7.5 Bures and quantum Hellinger metrics

7.5.1 Riemannian geometry on the manifold of invertible quantum states

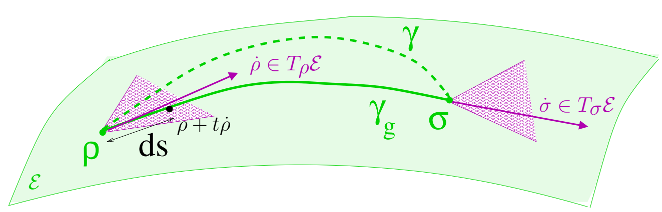

Figure 2: A shortest geodesic $\gamma_{\rm g}$ joining two states ρ and σ in the set of quantum states $\mathcal{E}(\mathcal{H})$ is a curve joining ρ and σ with minimal length.

In Riemannian geometry, a metric on a smooth manifold $\mathcal{X}$ is a smooth map g associating to each point $x \in \mathcal{X}$ a scalar product g_x on the tangent space $T_x\mathcal{X}$ at x. A curve γ on $\mathcal{X}$ joining two points x and y is parametrized by a piecewise C^1 map $\gamma : t \in [t_0,t_1] \mapsto \gamma(t) \in \mathcal{X}$ such that $\gamma(t_0) = x$ and $\gamma(t_1) = y$. Its length $\ell(\gamma)$ is

$$\ell(\gamma) = \int_\gamma \mathrm{d}s = \int_{t_0}^{t_1} \mathrm{d}t\,\sqrt{g_{\gamma(t)}(\dot\gamma(t),\dot\gamma(t))}\ , \tag{208}$$

where $\dot{\gamma}(t)$ stands for the time derivative $\mathrm{d}\gamma/\mathrm{d}t$. A Riemannian distance d on $\mathcal{X}$ can be associated to any metric g, defined as the infimum $d(x,y) = \inf_\gamma \ell(\gamma)$ of the lengths of all curves γ joining x and y. Such a distance is called the geodesic distance on $(\mathcal{X}, g)$. Curves $\gamma_{\rm g}$ with constant velocity which minimize the length *locally* are called geodesics[22]. In particular, if there is a geodesic $\gamma_{\rm g}$ with length $\ell(\gamma_{\rm g}) = d(x,y)$ minimizing the length globally, one says that $\gamma_{\rm g}$ is the shortest geodesic joining x and y (see Fig. 2).

Conversely, one can associate to a distance d on $\mathcal{X}$ a metric g if d satisfies the following condition (we ignore here the regularity assumptions): for any $x \in \mathcal{X}$ and $\dot{x} \in T_x\mathcal{X}$, the square distance between x and $x + t\dot{x}$ has the small time expansion[23]

$$\mathrm{d}s^2 = d(x, x+t\dot{x})^2 = g_\rho(\dot{x},\dot{x})t^2 + \mathcal{O}(t^3) \ . \tag{209}$$

The set of invertible quantum states on a finite dimensional Hilbert space $\mathcal{H}$ forms a smooth open manifold

$$\mathcal{X} = \mathcal{E}^{\,\rm inv}(\mathcal{H}) = \big\{ \rho \in \mathcal{B}(\mathcal{H})\ ;\ \rho > 0, \operatorname{tr}\rho = 1 \big\} \ . \tag{210}$$

Its tangent spaces $T_\rho\mathcal{E}(\mathcal{H})$ can be identified with the (real) vector space $\mathcal{B}(\mathcal{H})^0_{\rm s.a.}$ of self-adjoint traceless operators on $\mathcal{H}$. A trivial example of Riemannian distance on $\mathcal{E}^{\,\rm inv}(\mathcal{H})$ is the Hilbert-Schmidt distance d_2. Its metric is

$$(g_{\rm HS})_\rho(A,B) = \langle A, B\rangle = \operatorname{tr}(AB) \quad , \quad A, B \in \mathcal{B}(\mathcal{H})^0_{\rm s.a.} \ , \tag{211}$$

i.e., $(g_{\rm HS})_\rho$ is independent of ρ and given by the Hilbert-Schmidt scalar product. Introducing an orthonormal basis $\{|i\rangle\}_{i=1}^n$ of $\mathcal{H}$, one notes that $(g_{\rm HS})_\rho(A,B) = \sum_{i,j=1}^n \overline{A_{ij}} B_{ij}$ is nothing but the Euclidean scalar product. Thus the geodesics are straight lines,

$$\gamma_{\rm g}(\rho,\sigma) : t \in [t_0, t_1] \mapsto \frac{t_1 - t}{t_1 - t_0}\rho + \frac{t - t_0}{t_1 - t_0}\sigma \tag{212}$$

and one immediately checks that $\ell_{\rm HS}(\gamma_{\rm g}) = \operatorname{tr}[(-\rho+\sigma)^2]^{1/2} = d_2(\rho,\sigma)$. In contrast, the trace distance d_1 has not a well-defined metric.

In the remaining of this subsection, we show that the Bures and quantum Hellinger distances $d_{\rm B}$ and $d_{\rm H}$ have well-defined metrics $g_{\rm B}$ and $g_{\rm H}$. As we will see in subsection 7.8, the geodesic distance associated to the Bures metric $g_{\rm B}$ is not equal to $d_{\rm B}$ but to the so-called arccos Bures distance (see e.g. [134])

$$\widetilde{d_{\rm B}}(\rho,\sigma) = \inf_\gamma \ell_{\rm B}(\gamma) = \arccos\sqrt{F(\rho,\sigma)} = f(d_{\rm B}(\rho,\sigma)) \ , \tag{213}$$

where $f : [0,\sqrt{2}] \to [0,\pi/2]$ is the bijection $f(d) = \arccos(1 - d^2/2)$ with inverse $f^{-1}(\widetilde{d}) = 2\sin(\widetilde{d}/2)$. Note that $\widetilde{d_{\rm B}}$ and $d_{\rm B}$ have the same metrics because $f(d) = d + \mathcal{O}(d^3)$.

7.5.2 Bures metric

We start with the Bures distance. Let $\rho \in \mathcal{E}^{\,\rm inv}(\mathcal{H})$, $\dot{\rho} \in \mathcal{B}(\mathcal{H})^0_{\rm s.a.}$, and $t \in \mathbb{R}$, $|t| \ll 1$. According to Definition 7.2.1 one has

$$d_{\rm B}(\rho, \rho + t\dot{\rho})^2 = 2 - 2\operatorname{tr}(A(t)) \quad , \quad A(t) = \big(\sqrt{\rho}(\rho + t\dot{\rho})\sqrt{\rho}\big)^{\frac{1}{2}} \ . \tag{214}$$

[22] More precisely, $\gamma_{\rm g} : [t_0, t_1] \to \mathcal{X}$ is a geodesic if (i) $g_{\gamma_{\rm g}(t)}(\dot{\gamma_{\rm g}}(t), \dot{\gamma_{\rm g}}(t)) = $ const. and (ii) $\forall\ t \in (t_0, t_1)$, $\exists\ \delta > 0$ such that $\ell(\gamma_{\rm g}|_{[t,t+\delta]}) = d(\gamma_{\rm g}(t), \gamma_{\rm g}(t+\delta))$.

[23] Needless to say, determining the metric induced by a given Riemannian distance d is much simpler than finding an explicit formula for $d(x,y)$ given the metric g.

The scalar product $(g_{\rm B})_\rho$ will be given in terms of the eigenvectors $|k\rangle$ and eigenvalues p_k of ρ in the spectral decomposition $\rho = \sum_k p_k |k\rangle\langle k|$. Using the notation $\dot{A}(t) = {\rm d}A/{\rm d}t$, $\ddot{A}(t) = {\rm d}^2 A/{\rm d}t^2$ and deriving the identity $A(t)^2 = \sqrt{\rho}(\rho + t\dot{\rho})\sqrt{\rho}$ with respect to time, one finds

$$\begin{aligned} \dot{A}(0)A(0) + A(0)\dot{A}(0) &= \sqrt{\rho}\,\dot{\rho}\,\sqrt{\rho} \\ \ddot{A}(0)A(0) + 2\dot{A}(0)\dot{A}(0) + A(0)\ddot{A}(0) &= 0\,. \end{aligned} \tag{215}$$

By assumption, $A(0) = \rho$ is invertible. Since ${\rm tr}({\rm d}\rho) = 0$, it follows from the first equality in (215) that ${\rm tr}[\dot{A}(0)] = 0$. Multiplying the second equation in (215) by $A(0)^{-1}$ and taking the trace, one verifies that

$${\rm tr}[\ddot{A}(0)] = -\,{\rm tr}\big[\dot{A}(0)^2 A(0)^{-1}\big] = -\sum_{k,l=1}^{n} p_k^{-1} \big|\langle k|\dot{A}(0)|l\rangle\big|^2 = -\sum_{k,l=1}^{n} \frac{p_l |\langle k|{\rm d}\rho|l\rangle|^2}{(p_k+p_l)^2}\,. \tag{216}$$

Thus, going back to (214) we arrive at

$$d_{\rm B}(\rho, \rho + t{\rm d}\rho)^2 = -\,{\rm tr}[\ddot{A}(0)]t^2 + \mathcal{O}(t^3) = (g_{\rm B})_\rho({\rm d}\rho, {\rm d}\rho)t^2 + \mathcal{O}(t^3) \tag{217}$$

with (see [99, 166])

$$(g_{\rm B})_\rho(\dot{\rho},\dot{\rho}) = \frac{1}{2}\sum_{k,l=1}^{n} \frac{|\langle k|\dot{\rho}|l\rangle|^2}{p_k + p_l} \quad , \quad \dot{\rho} \in \mathcal{B}(\mathcal{H})^0_{\rm s.a.}\,,\ \rho > 0\,. \tag{218}$$

The last formula defines a scalar product on $\mathcal{B}(\mathcal{H})^0_{\rm s.a.}$ by polarization, hence $d_{\rm B}$ is Riemannian with metric $g_{\rm B}$. One readily obtains from this metric the infinitesimal volume element. The volume of $\mathcal{E}(\mathcal{H})$ and the area of its boundary have been determined in [166].

7.5.3 Quantum Hellinger metric

Let us now turn to the quantum Hellinger distance (199). We proceed to determine the metric g_α associated to the normal-ordered relative Rényi entropy (143), from which the quantum Hellinger metric $g_{\rm H}$ is obtained by setting $\alpha = 1/2$. We demonstrate that the largest metric g_α for $\alpha \in (0,1)$ is achieved for $\alpha = 1/2$ and is equal to $g_{\rm H}/2$, a result that will be needed later on (Sec. 8.1). The metric g_α is defined by

$$\begin{aligned} S_\alpha^{\rm (n)}(\rho + t{\rm d}\rho||\rho) &= (1-\alpha)^{-1}\big(1 - F_\alpha^{\rm (n)}(\rho + t{\rm d}\rho||\rho)^\alpha\big) + \mathcal{O}(t^3) \\ &= t^2(1-\alpha)^{-1}(g_\alpha)_\rho({\rm d}\rho, {\rm d}\rho) + \mathcal{O}(t^3)\,, \end{aligned} \tag{219}$$

where $F_\alpha^{\rm (n)}$ is the α-fidelity, see (200). To determine g_α for all $\alpha \in (0,1)$, we use (A1) in Appendix A to write

$$\begin{aligned} B_\alpha(t) &= \rho^\alpha - (\rho + t{\rm d}\rho)^\alpha = \frac{\sin(\alpha\pi)}{\pi}\int_0^\infty {\rm d}x\, x^\alpha \left(\frac{1}{x+\rho+t{\rm d}\rho} - \frac{1}{x+\rho}\right) \\ &= \frac{\sin(\alpha\pi)}{\pi}\int_0^\infty {\rm d}x\ x^\alpha \left(-\frac{t}{x+\rho}{\rm d}\rho\frac{1}{x+\rho} + \frac{t^2}{x+\rho}{\rm d}\rho\frac{1}{x+\rho}{\rm d}\rho\frac{1}{x+\rho}\right) + \mathcal{O}(t^3)\,. \end{aligned} \tag{220}$$

Introducing as before the spectral decomposition $\rho = \sum_k p_k |k\rangle\langle k|$ and using known integrals, one finds

$$\begin{aligned} 1 - F_\alpha^{\rm (n)}(\rho + t{\rm d}\rho||\rho)^\alpha &= {\rm tr}[B_\alpha(t)\rho^{1-\alpha}] \\ &= -t\alpha\sum_{k=1}^{n}\langle k|{\rm d}\rho|k\rangle + t^2\sum_{k,l=1}^{n}\frac{p_k^{1-\alpha}(p_k^\alpha - p_l^\alpha)}{(p_k-p_l)^2}\big|\langle k|{\rm d}\rho|l\rangle\big|^2 + \mathcal{O}(t^3)\,. \end{aligned} \tag{221}$$

Because $\mathrm{tr}(\mathrm{d}\rho) = 0$, the linear term in t vanishes as it should be. Plugging (221) into (219) one gets

$$(g_\alpha)_\rho(A, A) = \sum_{k,l=1}^{n} c_\alpha(p_k, p_l) |\langle k|A|l\rangle|^2 \quad , \quad c_\alpha(p, q) = \frac{(p^{1-\alpha} - q^{1-\alpha})(p^\alpha - q^\alpha)}{2(p - q)^2} \; . \tag{222}$$

It is easy to show that $c_\alpha(p, q) \leq c_{1/2}(p, q)$ for any $p, q > 0$, hence

$$\max_{\alpha \in (0,1)} (g_\alpha)_\rho(A, A) = \big(g_{1/2}\big)_\rho(A, A) = \sum_{k,l=1}^{n} \frac{|\langle k|A|l\rangle|^2}{2(\sqrt{p_k} + \sqrt{p_l})^2} \quad , \quad A \in \mathcal{B}(\mathcal{H})^0_{\text{s.a.}} \; , \tag{223}$$

as claimed above. Furthermore, in view of (199) we deduce that the quantum Hellinger distance d_{H} is Riemannian and has a metric $g_{\mathrm{H}} = 2g_{1/2}$. By comparing this metric with that of the Bures distance given by (218), one finds

$$(g_{\mathrm{B}})_\rho(A, A) \leq (g_{\mathrm{H}})_\rho(A, A) \leq 2(g_{\mathrm{B}})_\rho(A, A) \quad , \quad A \in \mathcal{B}(\mathcal{H})^0_{\text{s.a.}} \; , \; \rho > 0 \; . \tag{224}$$

These bounds implies the inequalities (201) between the Bures and quantum Hellinger distances in Proposition 7.4.1.

7.5.4 The qubit case

Consider a qubit with Hilbert space $\mathcal{H} \simeq \mathbb{C}^2$. Then the invertible density matrices $\rho \in \mathcal{E}^{\,\mathrm{inv}}(\mathbb{C}^2)$ are in one-to-one correspondance with vectors $\mathbf{r}$ in the open ball $B_1(\mathbb{R}^3) \subset \mathbb{R}^3$ of radius one and center 0 thanks to the Bloch representation [134]

$$\rho = \rho_{\mathbf{r}} = \frac{1}{2}(1 + \mathbf{r} \cdot \sigma) \; . \tag{225}$$

Here $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ is the vector formed by the three Pauli matrices and $\cdot$ stands for the standard scalar product in $\mathbb{R}^3$. The map $\mathbf{r} = (r_1, r_2, r_3) \in B_1(\mathbb{R}^3) \mapsto \rho_{\mathbf{r}}$ defines a (global) coordinate chart on the manifold $\mathcal{E}^{\,\mathrm{inv}}(\mathbb{C}^2)$. Any metric g on $\mathcal{E}^{\,\mathrm{inv}}(\mathbb{C}^2)$ is represented in these cartesian coordinates by ($\mathbf{r}$-dependent) 3×3 positive definite real matrices with elements $g_{ij}(\mathbf{r}) = g_{\rho_{\mathbf{r}}}(\partial_i \rho_{\mathbf{r}}, \partial_j \rho_{\mathbf{r}})$, where $\partial_i \rho_{\mathbf{r}} = \partial \rho_{\mathbf{r}} / \partial r_i = \sigma_i / 2$. Similarly, g is represented in spherical coordinates by the 3×3 positive definite matrices

$$\big(g_{\alpha\beta}(\mathbf{r})\big)_{\alpha,\beta = r,\theta,\varphi} = \big(g_{\rho_{\mathbf{r}}}(\partial_\alpha \rho_{\mathbf{r}}, \partial_\beta \rho_{\mathbf{r}})\big)_{\alpha,\beta = r,\theta,\varphi} = J^T \,\big(g_{ij}(\mathbf{r})\big)_{i,j=1}^{3} \, J \; , \tag{226}$$

where $\partial_r \rho_{\mathbf{r}} = \partial \rho_{\mathbf{r}} / \partial r$ and similarly for $\partial_\theta \rho_{\mathbf{r}}$ and $\partial_\varphi \rho_{\mathbf{r}}$. In the last expression in (226), J is the Jacobian matrix of the transformation from cartesian to spherical coordinates.

It is an instructive exercise to determine the forms in cartesian and spherical coordinates of the Bures and Hellinger metrics for a qubit, by applying the general formulas (218) and (223). The eigenvalues and normalized eigenvectors of $\rho_{\mathbf{r}}$ are

$$p_\pm = \frac{1 \pm |\mathbf{r}|}{2} \quad , \quad |\pm\rangle = \frac{1}{\sqrt{2|\mathbf{r}|(|\mathbf{r}| \pm r_3)}} \begin{pmatrix} r_3 \pm |\mathbf{r}| \\ r_1 + \mathrm{i} r_2 \end{pmatrix} \; . \tag{227}$$

In particular, $\rho_{\mathbf{r}} > 0$ if and only if $r = |\mathbf{r}| < 1$ and the pure states are associated to vectors on the Bloch sphere $\partial B_1(\mathbb{R}^3)$. An explicit calculation gives

$$\langle \pm|\sigma_i|\pm\rangle = \pm \frac{r_i}{|\mathbf{r}|} \quad , \quad \mathrm{Re}\left\{\langle +|\sigma_i|-\rangle \overline{\langle +|\sigma_j|-\rangle}\right\} = \delta_{ij} - \frac{r_i r_j}{|\mathbf{r}|^2} \quad i, j = 1, 2, 3 \; , \tag{228}$$

with δ_{ij} the Kronecker symbol. By using (218) and (223) and the polarization identity to obtain $g_\rho(\sigma_i,\sigma_j)$ for $i \neq j$, one deduces that the Bures and Hellinger metrics are represented in the cartesian coordinate basis by

$$\begin{aligned}
(g_{\rm B})_{ij}(\mathbf{r}) &= \frac{1}{4}\biggl(\frac{r_i r_j}{1-|\mathbf{r}|^2} + \delta_{ij}\biggr) \\
(g_{\rm H})_{ij}(\mathbf{r}) &= \frac{1}{4}\biggl(\frac{1}{1-|\mathbf{r}|^2} - \frac{2}{1+\sqrt{1-|\mathbf{r}|^2}}\biggr)\frac{r_i r_j}{|\mathbf{r}|^2} + \frac{\delta_{ij}}{2(1+\sqrt{1-|\mathbf{r}|^2})}\,.
\end{aligned} \tag{229}$$

To obtain the matrices in spherical coordinates, we note that for any $\alpha,\beta = r,\theta,\varphi$ one has

$$\sum_{i,j=1}^{3} J_{i\alpha} r_i r_j J_{j\beta} = \frac{1}{4}\partial_\alpha(|\mathbf{r}|^2)\partial_\beta(|\mathbf{r}|^2) = r^2 \delta_{\alpha\beta}\delta_{\alpha r} \tag{230}$$

$$\sum_{i,j=1}^{3} J_{i\alpha}\delta_{ij} J_{j\beta} = \partial_\alpha \mathbf{r}\cdot\partial_\beta\mathbf{r} = \delta_{\alpha\beta}\bigl(\delta_{\alpha r} + r^2\,\delta_{\alpha\theta} + r^2\sin^2\theta\,\delta_{\alpha\varphi}\bigr)\,. \tag{231}$$

Thus the matrices of $g_{\rm B}$ and $g_{\rm H}$ in the spherical coordinate basis are diagonal and given respectively by

$$\bigl(g_{\alpha\beta}(\mathbf{r})\bigr)_{\alpha,\beta=r,\theta,\varphi} = \frac{1}{4}\begin{pmatrix} \frac{1}{1-r^2} & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2\theta \end{pmatrix} \quad , \quad \frac{1}{4}\begin{pmatrix} \frac{1}{1-r^2} & 0 & 0 \\ 0 & \frac{2r^2}{1+\sqrt{1-r^2}} & 0 \\ 0 & 0 & \frac{2r^2\sin^2\theta}{1+\sqrt{1-r^2}} \end{pmatrix} \quad , \quad 0 \leq r < 1\,. \tag{232}$$

The corresponding square infinitesimal distance elements are

$$\mathrm{d}s_{\rm B}^2 = \frac{1}{4}\biggl(\frac{\mathrm{d}r^2}{1-r^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\varphi^2\biggr) \ , \ \mathrm{d}s_{\rm H}^2 = \frac{1}{4}\biggl(\frac{\mathrm{d}r^2}{1-r^2} + \frac{2r^2(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\varphi^2)}{1+\sqrt{1-r^2}}\biggr)\,. \tag{233}$$

We observe that in both cases the matrix element $g_{rr}(\mathbf{r})$ diverges when $r \to 1$, that is, when one approaches the boundary of $\mathcal{E}(\mathbb{C}^2)$ transversally. This singularity can be removed by changing coordinates, defining $t = \arcsin(r)$. The square infinitesimal distance elements become

$$\mathrm{d}s_{\rm B}^2 = \frac{1}{4}\Bigl(\mathrm{d}t^2 + \sin^2 t\,\bigl(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\bigr)\Bigr) \ , \ \mathrm{d}s_{\rm H}^2 = \frac{1}{4}\Bigl(\mathrm{d}t^2 + \frac{2\sin^2 t}{1+\cos t}\bigl(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\varphi^2\bigr)\Bigr)\,, \tag{234}$$

where $t \in [0,\pi/2]$, $\theta \in [0,\pi]$, and $\varphi \in [0,2\pi]$. One can associate to (t,θ,φ) a point $\mathbf{x}$ on the upper half of the hypersphere S^3 of unit radius in $\mathbb{R}^4$, with coordinates

$$\left\{\begin{array}{lllll} x_1 &=& r_1 &=& \sin t\,\sin\theta\,\cos\varphi \\ x_2 &=& r_2 &=& \sin t\,\sin\theta\,\cos\varphi \\ x_3 &=& r_3 &=& \sin t\,\cos\theta \\ x_4 &=& \cos t && \end{array}\right. . \tag{235}$$

Hence qubit states are in one-to-one correspondance with points $\mathbf{x}$ on the upper half hypersphere $S^3_+ \subset \mathbb{R}^4$. The qubit density matrix ρ corresponding to $\mathbf{x} \in S^3_+$ is obtained from (225) where $\mathbf{r} \in B_1(\mathbb{R}^3)$ is the projection of $\mathbf{x}$ perpendiculary to the vertical axis x_4. The equator of S^3, which can be identified with a two-dimensional unit sphere S^2, corresponds to pure states (angle $t = \pi/2$).

It is worth noting that the Bures metric in the (t,θ,φ)-coordinates coincides up to a factor of $1/4$ with the metric of the hypersphere S^3. Therefore $\mathcal{E}(\mathbb{C}^2)$ equipped with the Bures metric can be identified with S^3_+. This implies in particular that $(\mathcal{E}(\mathbb{C}^2), g_{\rm B})$ has a constant positive curvature and that the geodesics are projections of big circles of S^3_+ perpendicularly to the x_4-axis (see Sec. 7.7.7 below). The singularity of the metric in the $\mathbf{r}$-coordinates as $r \to 1$ is a by-product of this projection.

7.6 Quantum Fisher information and skew information

In this section we introduce two quantities related to the Bures and Hellinger metrics g_{B} and g_{H}, the quantum Fisher information and the skew information, which play an important role in quantum metrology.

7.6.1 Quantum Fisher information for unitary transformations

Definition 7.6.1. *Given a state $\rho \in \mathcal{E}(\mathcal{H})$ and an observable $H \in \mathcal{B}(\mathcal{H})_{\mathrm{s.a.}}$, the quantum Fisher information of ρ with respect to H is the non-negative number $\mathcal{F}_Q(\rho, H)$ such that*

$$d_B(\rho, \rho_\varepsilon)^2 = \frac{\varepsilon^2}{4}\mathcal{F}_Q(\rho, H) + \mathcal{O}(\varepsilon^3) \tag{236}$$

with $\rho_\theta = e^{-\mathrm{i}\theta H}\rho\, e^{\mathrm{i}\theta H}$. If ρ is invertible, then $\mathcal{F}_Q(\rho, H) = 4(g_{\mathrm{B}})_\rho\big(-\mathrm{i}[H,\rho], -\mathrm{i}[H,\rho]\big)$.

Thus the square root of the quantum Fisher information gives (up to a factor of two) the "speed" at which a given state ρ separates from its θ-evolved state ρ_θ under the dynamics specified by the Hamiltonian H,

$$\sqrt{\mathcal{F}_Q(\rho, H)} = 2\frac{\mathrm{d}}{\mathrm{d}\theta} d_{\mathrm{B}}(\rho, \rho_\theta)\Big|_{\theta=0} . \tag{237}$$

Note that $\mathcal{F}_Q(\rho_\theta, H) = \mathcal{F}_Q(\rho, H)$ by unitary invariance of the Bures distance. The quantity $\mathcal{F}_Q(\rho, H)$ has been introduced by different authors [83, 84, 92, 33] as a quantum analog of the Fisher information in statistics[24] . Similarly to the definition of the Bures distance in Sec. 7.2, Braunstein and Caves relate it to the metric – called the "distinguishability metric" by Wootters [194] – extending the Fubini-Study metric to mixed states. For a pure state $\rho_\psi = |\psi\rangle\langle\psi|$, the quantum Fisher information reduces to the square quantum fluctuation of H, namely,

$$\mathcal{F}_Q(\rho_\psi, H) = 4\langle(\Delta H)^2\rangle_\psi = 4\big(\langle\psi|H^2|\psi\rangle - \langle\psi|H|\psi\rangle^2\big) . \tag{238}$$

This can be proven either from an explicit calculation of the Fubini-Study distance in (236) or by using Proposition 7.6.2 below.

The quantum Fisher information plays an important role in quantum metrology. As we will see in Sec. 8.2, it quantifies the amount of information on an unknown real parameter θ contained in the states ρ_θ when the system undergoes the θ-dependent unitary transformation $\rho \to \rho_\theta = e^{-\mathrm{i}\theta H}\rho\, e^{\mathrm{i}\theta H}$ with Hamiltonian H. Such information is extracted from ρ_θ by performing an optimal measurement on the system and using an optimal estimator of θ depending on the measurement outcomes. It will be argued below that $\mathcal{F}_Q(\rho, H)$ quantifies the amount of quantum correlations in the input state ρ enabling to improve precision in the estimation of θ beyond the shot-noise limit (see Sec. 8.2).

We next prove two nice properties of $\mathcal{F}_Q(\rho, H)$.

---

[24] In these references, the quantum Fisher information is defined by the formula

$$\mathcal{F}_Q(\rho, H) = \operatorname{tr}\big[\rho\,(\mathcal{K}_{\mathrm{B}})_\rho^{-1}(\dot\rho)(\mathcal{K}_{\mathrm{B}})_\rho^{-1}(\dot\rho)\big] = \operatorname{tr}\big[\dot\rho\,(\mathcal{K}_{\mathrm{B}})_\rho^{-1}(\dot\rho)\big] , \tag{$*$}$$

where $\dot\rho = -\mathrm{i}[H\,,\,\rho]$ and the symmetric logarithmic derivative $(\mathcal{K}_{\mathrm{B}})_\rho^{-1}$ is the inverse of the operator $(\mathcal{K}_{\mathrm{B}})_\rho = \frac{1}{2}(\mathcal{L}_\rho + \mathcal{R}_\rho)$ on $\mathcal{B}(\mathcal{H})$ with $\mathcal{L}_\rho$ and $\mathcal{R}_\rho$ defined in (7), i.e.,

$$(\mathcal{K}_{\mathrm{B}})_\rho(A) = \frac{1}{2}\big(\rho A + A\rho\big) \quad , \quad A \in \mathcal{B}(\mathcal{H}) .$$

The last equality in ($*$) follows from the cyclicity of the trace. It is easy to show that the last trace in ($*$) equals the expression in the right-hand side of the identity (239) in Proposition 7.6.2.

Proposition 7.6.2. *For any $\rho \in \mathcal{E}(\mathcal{H})$ (invertible or not), one has*

$$\mathcal{F}_Q(\rho, H) = 2 \sum_{k,l,p_k+p_l>0} \frac{(p_k - p_l)^2}{p_k + p_l} |\langle k|H|l\rangle|^2 \,, \tag{239}$$

where $\{|k\rangle\}$ is an orthonormal basis of eigenvectors of ρ with eigenvalues p_k.

Proof. The result follows from formula (218) for the Bures metric in the case of invertible ρ. We now repeat the steps of the derivation of this formula for non-invertible ρ. It is necessary here to consider the terms of order ε^2 in the expansion of $\rho_\varepsilon = e^{-\mathrm{i}\,\varepsilon H} \rho \, e^{\mathrm{i}\,\varepsilon H}$, yielding

$$A(\varepsilon)^2 = \sqrt{\rho}\,\rho_\varepsilon \sqrt{\rho} = \sqrt{\rho}\Big(\rho - \mathrm{i}\,\varepsilon\big[H, \rho\big] + \varepsilon^2 \Big(H\rho H - \frac{1}{2}\big\{H^2, \rho\big\}\Big) + \mathcal{O}(\varepsilon^3)\Big)\sqrt{\rho} \,. \tag{240}$$

In fact, the terms $\mathcal{O}(\varepsilon^2)$ contribute to the second derivative $\ddot{A}(0)$ of $A(\varepsilon)$. Equations (215) become

$$\begin{aligned} \dot{A}(0)A(0) + A(0)\dot{A}(0) &= -\mathrm{i}\sqrt{\rho}\,\big[H, \rho\big]\,\sqrt{\rho} \\ \ddot{A}(0)A(0) + 2\dot{A}(0)\dot{A}(0) + A(0)\ddot{A}(0) &= 2\sqrt{\rho}\Big(H\rho H - \frac{1}{2}\big\{H^2, \rho\big\}\Big)\sqrt{\rho} \,. \end{aligned} \tag{241}$$

Let r be the rank of ρ. We denote by P the projector onto $\operatorname{supp}(\rho) = \operatorname{span}\{|k\rangle\,;\, k = 1, \ldots, r\}$ and by $p_1, \ldots, p_r > 0$ the positive eigenvalues of ρ. One deduces from the identity $A(t) = PA(t)P$ that $\dot{A}(0) = P\dot{A}(0)P$ and $\ddot{A}(0) = P\ddot{A}(0)P$. Taking the matrix elements of the operator identities (241) in the eigenbasis of $\rho = A(0)$ one infers that, for any $k, l = 1, \ldots r$,

$$\begin{aligned} (p_k + p_l)\langle k|\dot{A}(0)|l\rangle &= -\mathrm{i}\sqrt{p_k p_l}(p_l - p_k)\langle k|H|l\rangle \\ p_k\langle k|\ddot{A}(0)|k\rangle &= -\langle k|(P\dot{A}(0)P)^2|k\rangle + p_k\langle k|H\rho H|k\rangle - p_k^2\langle k|H^2|k\rangle \\ &= -\sum_{l=1}^{r} \big|\langle k|\dot{A}(0)|l\rangle\big|^2 + p_k \sum_{l=1}^{r} p_l \big|\langle k|H|l\rangle\big|^2 - p_k^2\langle k|H^2|k\rangle \,. \end{aligned} \tag{242}$$

By combining these two equations one is led to

$$\operatorname{tr}[\ddot{A}(0)] = \operatorname{tr}[P\ddot{A}(0)P] = \sum_{k=1}^{r}\langle k|\ddot{A}(0)|k\rangle = -\sum_{k,l=1}^{r} \frac{p_l(p_l - p_k)^2|\langle k|H|l\rangle|^2}{(p_k + p_l)^2} - \sum_{l=1}^{r} p_l\langle l|H(1-P)H|l\rangle \,. \tag{243}$$

Moreover, from the first identity in (242) one has $\operatorname{tr}[\dot{A}(0)] = \operatorname{tr}[P\dot{A}(0)P] = 0$. Thus

$$d_B\big(\rho, \rho_\varepsilon\big)^2 = 2 - 2\operatorname{tr}[A(\varepsilon)] = -\varepsilon^2 \operatorname{tr}[\ddot{A}(0)] + \mathcal{O}(\varepsilon^3) \,. \tag{244}$$

The result follows by using the symmetrization trick in the double sum in (243) and plugging this equation into (244). □

Proposition 7.6.3. (Quantum Fisher information as a convex roof) [174, 198]). *The quantum Fisher information $\mathcal{F}_Q(\rho, H)$ is convex in ρ and is the convex roof of the variance, i.e.*

$$\mathcal{F}_Q(\rho, H) = 4 \min_{\{|\phi_j\rangle, \zeta_j\}} \sum_{j=1}^{p} \zeta_j \langle(\Delta H)^2\rangle_{\phi_j} \,, \tag{245}$$

where the infimum is over all pure state decompositions $\{|\phi_j\rangle, \zeta_j\}_{j=1}^{p}$ of ρ with $p \le \dim(\mathcal{H})$.

It is shown in [174, 198] that the maximum of the averaged variance $\sum_j \zeta_i \langle (\Delta H)^2 \rangle_{\phi_j}$ over all pure state decompositions of ρ is, in turn, given by the usual variance

$$\langle (\Delta H)^2 \rangle_\rho = \mathrm{tr}[\rho H^2] - (\mathrm{tr}[\rho H])^2 \ . \tag{246}$$

Hence one has

$$\frac{1}{4}\mathcal{F}_Q(\rho, H) \leq \sum_{j=1}^{p} \zeta_j \langle (\Delta H)^2 \rangle_{\phi_j} \leq \langle (\Delta H)^2 \rangle_\rho \tag{247}$$

and there exist pure state decompositions for which the inequalities are equalities. The first bound in (247) results from the convexity of $\mathcal{F}_Q(\rho, H)$ claimed in the Proposition and the fact that the quantum Fisher information of a pure state is equal to four times the variance of H in this state, see (238). The second bound follows from the concavity of $\langle (\Delta H)^2 \rangle_\rho$. The convex roof formula (245) was first proven in [174] for states ρ of rank two and conjectured to hold in general, and was subsequently proven in [198] using a tricky guess for the optimal pure state decomposition. We present below an alternative proof based on Uhlmann's theorem (Theorem 7.2.2) and the solution of an optimization problem for quantum expectations of product of observables in a bipartite system (Lemma 7.6.4).

Proof of Proposition 7.6.3. The convexity of $\mathcal{F}_Q(\rho, H)$ follows from the definition (236) and the joint convexity of $d_{\mathrm{B}}(\rho, \sigma)^2$ (Corollary 7.2.3). As explained above, it implies the first bound in (247). It remains to show that this bound is tight. From (236) and the definition (178) of the Bures distance, one has

$$1 - F(\rho, \rho_\varepsilon) = \frac{\varepsilon^2}{4}\mathcal{F}_Q(\rho, H) + \mathcal{O}(\varepsilon^3) \quad , \quad \rho_\varepsilon = e^{-\mathrm{i}\,\varepsilon H} \rho\, e^{\mathrm{i}\,\varepsilon H} \ . \tag{248}$$

For any ε in a neighborhood of 0, let

$$|\Phi_\varepsilon\rangle = e^{-\mathrm{i}\varepsilon H \otimes 1} |\Phi_0\rangle \tag{249}$$

be a (fixed) purification of ρ_ε on $\mathcal{H} \otimes \mathcal{K}$. We first apply Theorem 7.2.2 to express the fidelity $F(\rho, \rho_\varepsilon)$ as a maximum over all purifications $|\Psi\rangle$ of ρ on $\mathcal{H} \otimes \mathcal{K}$,

$$F(\rho, \rho_\varepsilon) = \max_{|\Psi\rangle} |\langle \Psi | \Phi_\varepsilon \rangle|^2 \ . \tag{250}$$

Recall that the ancilla space $\mathcal{K}$ can be chosen of dimension $n = \dim \mathcal{H}$. By expanding $|\Phi_\varepsilon\rangle$ for small ε, one gets

$$|\langle \Psi | \Phi_\varepsilon \rangle|^2 = |\langle \Psi | \Phi_0 \rangle|^2 + 2\varepsilon \,\mathrm{Re}\, \langle \Psi | \Phi_0 \rangle \langle \dot{\Phi}_0 | \Psi \rangle + \varepsilon^2 \Big(\mathrm{Re}\, \langle \Psi | \Phi_0 \rangle \langle \ddot{\Phi}_0 | \Psi \rangle + |\langle \Psi | \dot{\Phi}_0 \rangle|^2 \Big) + \mathcal{O}(\varepsilon^3) \tag{251}$$

with $|\dot{\Phi}_0\rangle = -\mathrm{i} H \otimes 1 |\Phi_0\rangle$ and $|\ddot{\Phi}_0\rangle$ the first and second derivatives of $|\Phi_\theta\rangle$ at $\theta = 0$. Recall that all purifications of ρ on $\mathcal{H} \otimes \mathcal{K}$ are given by $|\Psi\rangle = 1 \otimes V |\Phi_0\rangle$ with $|\Phi_0\rangle$ a fixed purification of ρ and V a unitary operator acting on $\mathcal{K}$. The purification of ρ maximizing the fidelity (251) depends on ε, we will denote it by $|\Psi^{(\varepsilon)}\rangle$. It is such that $|\Psi^{(\varepsilon)}\rangle = |\Phi_0\rangle + \mathcal{O}(\varepsilon)$ for $\varepsilon \ll 1$, up to an irrelevant phase factor (the maximal fidelity is then equal to $1 + \mathcal{O}(\varepsilon)$). Thus this optimal purification has the form $|\Psi^{(\varepsilon)}\rangle = 1 \otimes e^{-\mathrm{i}(\varepsilon B + \varepsilon^2 C + \mathcal{O}(\varepsilon^3))} |\Phi_0\rangle$ for some ε-independent self-adjoint operators B and C on $\mathcal{K}$. Plugging this value into (251), expanding in powers of ε and noting that $\mathrm{Re}\, \langle \dot{\Phi}_0 | \Phi_0 \rangle = 0$ and $\mathrm{Re}\, \langle \ddot{\Phi}_0 | \Phi_0 \rangle = -\| \dot{\Phi}_0 \|^2$, one finds

$$|\langle \Psi^{(\varepsilon)} | \Phi_\varepsilon \rangle|^2 = 1 + \varepsilon^2 \Big(-\langle 1 \otimes (\Delta B)^2_{\Phi_0} \rangle_{\Phi_0} + 2\, \mathrm{Im}\, \langle \dot{\Phi}_0 | 1 \otimes (\Delta B)_{\Phi_0} | \Phi_0 \rangle - \| \dot{\Phi}_0 \|^2 + |\langle \Phi_0 | \dot{\Phi}_0 \rangle|^2 \Big) + \mathcal{O}(\varepsilon^3) \tag{252}$$

with $(\Delta B)_{\Phi_0} = B - \langle 1 \otimes B \rangle_{\Phi_0}$. Note that C does not contribute to the pure state fidelity to second order in ε. The fidelity $F(\rho, \rho_\varepsilon)$ equals the maximum of this expression over all self-adjoint operators B on $\mathcal{K}$. By replacing (252) into (248) one obtains

$$\mathcal{F}_Q(\rho_0, H) = 4 \min_{B \in \mathcal{B}(\mathcal{K})_{\rm s.a.}} \left\langle \Big((\Delta H)_{\Phi_0} \otimes 1 - 1 \otimes (\Delta B)_{\Phi_0} \Big)^2 \right\rangle_{\Phi_0} . \tag{253}$$

The desired result then follows from the following lemma.

Lemma 7.6.4. *Let $\rho \in \mathcal{E}(\mathcal{H})$ and $|\Psi\rangle$ be a purification of ρ on $\mathcal{H} \otimes \mathcal{K}$. Given two local observables A and B on $\mathcal{H}$ and $\mathcal{K}$, we write $(\Delta A)_\Psi = A - \langle A \otimes 1 \rangle_\Psi = A - \langle A \rangle_\rho$ and $(\Delta B)_\Psi = B - \langle 1 \otimes B \rangle_\Psi$. Then*

$$\min_{B \in \mathcal{B}(\mathcal{K})_{\rm s.a.}} \left\langle \Big((\Delta A)_\Psi \otimes 1 - 1 \otimes (\Delta B)_\Psi \Big)^2 \right\rangle_\Psi = \min_{\{|\phi_j\rangle, \zeta_j\}} \sum_{j=1}^{p} \zeta_j \big(\langle A^2 \rangle_{\phi_j} - \langle A \rangle_{\phi_j}^2 \big) , \tag{254}$$

where the minimum is over all pure state decompositions $\{|\phi_j\rangle, \zeta_j\}$ of ρ with $p \leq \dim \mathcal{K}$ states.

Proof. Let us first observe that the left-hand side of (254) does not depend upon the choice of the purification $|\Psi\rangle$, as one easily convinces oneself by using that different purifications are related by a unitary operator acting on $\mathcal{K}$. Let us write $|\Psi\rangle = \sum_i \sqrt{\eta_i} |\psi_i\rangle |i\rangle$ with $\{|i\rangle\}_{i=1}^m$ an orthonormal basis of $\mathcal{K}$. Recall that one can associate to the purification $|\Psi\rangle$ a pure state decomposition $\{|\psi_i\rangle, \eta_i\}_{i=1}^m$ of ρ. Furthermore, any other pure state decomposition $\{|\phi_j\rangle, \zeta_j\}_{j=1}^p$ with $p \leq m$ is related to the former decomposition by (16), where $(u_{ji})_{i,j=1}^m$ is a $m \times m$ unitary matrix. Next, note that any operator $B \in \mathcal{B}(\mathcal{K})_{\rm s.a.}$ can be written as $B = U^* D U$ with $D = \sum_j b_j |j\rangle\langle j|$ diagonal and $U \in \mathcal{B}(\mathcal{K})$ unitary. Therefore,

$$\begin{aligned} \big\langle 1 \otimes (\Delta B)_\Psi^2 \big\rangle_\Psi &= \sum_{i,j,k=1}^{m} \sqrt{\eta_i \eta_k} \langle \psi_i | \psi_k \rangle \overline{u_{ji}}\, b_j^2 u_{jk} - \Bigg(\sum_{i,j,k=1}^{m} \sqrt{\eta_i \eta_k} \langle \psi_i | \psi_k \rangle \overline{u_{ji}}\, b_j u_{jk} \Bigg)^2 \\ &= \sum_{j=1}^{p} \zeta_j b_j^2 - \Bigg(\sum_{j=1}^{p} \zeta_j b_j \Bigg)^2 = \sum_{j=1}^{p} \zeta_j (\Delta b)_j^2 \end{aligned} \tag{255}$$

with $(\Delta b)_j = b_j - \sum_k \zeta_k b_k = b_j - \langle 1 \otimes B \rangle_\Psi$. Similarly,

$$\big\langle (\Delta A)_\Psi \otimes (\Delta B)_\Psi \big\rangle_\Psi = \sum_{j=1}^{p} \zeta_j (\Delta a)_j (\Delta b)_j \quad , \quad (\Delta a)_j = \langle A \rangle_{\phi_j} - \langle A \rangle_\rho . \tag{256}$$

We may rewrite the minimum over all self-adjoint operators B as a minimum over all vectors $\mathbf{b} \in \mathbb{R}^m$ (eigenvalues of B) and all unitaries U or, equivalently, all pure state decompositions of ρ with $p \leq m$ states. This gives

$$\begin{aligned} \min_{B \in \mathcal{B}(\mathcal{K})_{\rm s.a.}} \Big\langle -2 (\Delta A)_\Psi \otimes (\Delta B)_\Psi + 1 \otimes (\Delta B)_\Psi^2 \Big\rangle_\Psi &= \min_{\{|\phi_j\rangle, \zeta_j\}} \min_{\mathbf{b}} \sum_{j=1}^{p} \zeta_j \Big(-2 (\Delta a)_j (\Delta b)_j + (\Delta b)_j^2 \Big) \\ &= \min_{\{|\phi_j\rangle, \zeta_j\}} \Bigg\{ - \sum_{j=1}^{p} \zeta_j (\Delta a)_j^2 \Bigg\} . \end{aligned} \tag{257}$$

Formula (254) follows by observing that $\langle A \rangle_\rho = \sum_j \zeta_j \langle A \rangle_{\phi_j}$ and $\langle (\Delta A)_\Psi^2 \otimes 1 \rangle_\Psi = \sum_j \zeta_j \langle A^2 \rangle_{\phi_j} - \langle A \rangle_\rho^2$ (note that these quantities are independent of the chosen purification $\{|\phi_j\rangle, \zeta_j\}$). $\square$

7.6.2 Quantum Fisher information for non-unitary transformations

When studying parameter estimation in quantum metrology in the presence of noise, the output states are related to the input state through a quantum operation depending on the parameter θ to be estimated (for more details see Sec. 8.2). The corresponding quantum Fisher information is given by the following generalization of definition 7.6.1 to non-unitary transformations. Given a $C^2(I)$-family $\{\mathcal{M}_\theta\}_{\theta\in I}$ of quantum operations $\mathcal{M}_\theta$ on $\mathcal{H}$ depending on a real parameter $\theta \in I$ and an input state $\rho_{\rm in}$, one defines $\mathcal{F}_Q(\rho_{\theta_0}, \{\mathcal{M}_\theta\}_\theta)$ as in formula (236) by

$$d_B(\rho_{\theta_0}, \rho_\theta)^2 = \frac{(\theta-\theta_0)^2}{4}\mathcal{F}_Q(\rho_{\theta_0}, \{\mathcal{M}_\theta\}_\theta) + \mathcal{O}\big((\theta-\theta_0)^3\big) , \tag{258}$$

where $\rho_\theta = \mathcal{M}_\theta(\rho_{\rm in})$. In contrast to the unitary case, this Fisher information depends on the parameter value θ_0 around which ρ_θ is expanded. For $\rho > 0$ this yields $4(g_{\rm B})_\rho(\dot\rho, \dot\rho)$, where $\dot\rho$ is the first derivative of ρ_θ at θ_0.

By repeating the arguments in the proof of Proposition 7.6.2, replacing the right-hand side of (240) by $\sqrt{\rho}(\rho + \varepsilon\dot\rho + \varepsilon^2\ddot\rho/2 + \mathcal{O}(\varepsilon^2))\sqrt{\rho}$ with $\varepsilon = \theta - \theta_0$, one finds[25]

$$\mathcal{F}_Q(\rho_{\theta_0}, \{\mathcal{M}_\theta\}_{\theta\in I}) = 4\bigg(\frac{\mathrm{d}}{\mathrm{d}\theta} d_{\rm B}(\rho_{\theta_0}, \rho_\theta)\Big|_{\theta=\theta_0}\bigg)^2 = 2\sum_{k,l,p_k+p_l>0} \frac{|\langle k|\dot\rho|l\rangle|^2}{p_k+p_l} + 2\sum_{k,p_k=0} \ddot p_k , \tag{259}$$

where $\{|k\rangle\}$ is an orthonormal basis of eigenvectors of ρ_{θ_0} with eigenvalues p_k. For pure state evolutions, one easily finds

$$\mathcal{F}_Q(|\psi_{\theta_0}\rangle, \{|\psi_\theta\rangle\}_{\theta\in I}) = 4\big(\|\dot\psi_{\theta_0}\|^2 - \big|\langle\dot\psi_{\theta_0}|\psi_{\theta_0}\rangle\big|^2\big) . \tag{260}$$

Note the last sum in (259) involving the non-negative second derivatives $\ddot p_k = \mathrm{d}^2 p_k/\mathrm{d}\theta^2|_{\theta=\theta_0}$ of the eigenvalues $p_k(\theta)$ of ρ_θ vanishing at $\theta=\theta_0$. The first sum in (259) corresponds to the quantum Fisher information usually defined in the literature, given in terms of a symmetric logarithmic derivative as in formula $(*)$ in the footnote after definition 7.6.1. The two definitions coincide for unitary evolutions and, more generally, whenever $\mathrm{rank}(\rho_\theta)$ is continuous at θ_0. In the opposite case, i.e., when ρ_θ has at least one eigenvalue $p_k(\theta)$ vanishing at $\theta=\theta_0$ and nonzero for $\theta<\theta_0$ or $\theta>\theta_0$, the definition in terms of the symmetric logarithmic derivatives has the drawback to yield to discontinuities as ρ_{θ_0} moves along the trajectory $\theta\mapsto\rho_\theta$ [155]. In contrast, as a consequence of the smoothness of the distance, our definition (258) yields a continuous Fisher information (see [155] for an explicit proof). This can be illustrated by the following example. Assume that $\{\rho_\theta = \sum_k p_k(\theta)|k\rangle\langle k|\}_{\theta\in I}$ is a family of commuting states, with $\{|k\rangle\}$ a fixed basis. Then the first sum in (259) equals the classical Fisher information $\sum_k \dot p_k(\theta_0)^2/p_k(\theta_0)$ and has a jump for trajectories bouncing on the boundary of the set of quantum states at $\theta=\theta_0$. More precisely, each eigenvalue with a minimum $p_k(\theta_0) = \dot p_k(\theta_0) = 0$ at θ_0 contributes to the jump by an amplitude given by $\lim_{\theta\to\theta_0} \dot p_k(\theta)^2/p_k(\theta) = 2\ddot p_k(\theta_0)$. On the other hand, $(g_{\rm B})_{\rho_\theta}(\dot\rho(\theta), \dot\rho(\theta))$ is continuous at θ_0 due to the aforementioned additional term coanceling the discontinuity.

---

[25] Here one uses the fact if ρ_θ has an eigenvalue $p_k(\theta)$ vanishing at $\theta=\theta_0$ then the first and second derivatives of $p_k(\theta)$ at $\theta=\theta_0$ satisfy $\dot p_k = 0$ and $\ddot p_k \geq 0$, because $p_k(\theta)\geq 0$ for all $\theta\in I$ (since ρ_θ is a quantum state). As a result, $\langle k|\dot\rho|k\rangle = 0$ if $p_k = 0$. One can then show from the analog of (242) that $2\,\mathrm{tr}[\dot A(0)] = \mathrm{tr}[P\dot\rho] = \mathrm{tr}[\dot\rho] = 0$ with P the projector onto $\mathrm{supp}(\rho)$. The second sum in the right-hand side of (259) appears by using the following identity, which also follows from the aforementioned fact and from $\mathrm{tr}[\ddot\rho]=0$:

$$\mathrm{tr}[P\ddot\rho] = -\,\mathrm{tr}[(1-P)\ddot\rho(1-P)] = -2\sum_{p_k>0,p_l=0}\frac{|\langle k|\dot\rho|l\rangle|^2}{p_k} - \sum_{l,p_l=0}\ddot p_l .$$

A variational formula for $\mathcal{F}_Q(\rho_{\theta_0}, \{\mathcal{M}_\theta\}_{\theta\in I})$ (with the definition (258)) has been derived in [63]. Let $\{|\Phi_\theta\rangle\}_{\theta\in I}$ be a (fixed) family of purifications of $\rho_\theta = \mathcal{M}_\theta(\rho_{\rm in})$, for instance

$$|\Phi_\theta\rangle = U_\theta \otimes 1_{\sf A}|\psi_{\rm in}\rangle|\epsilon_0\rangle \ \in \ \mathcal{H}_{\sf SEA} = \mathcal{H}_{\sf S}\otimes\mathcal{H}_{\sf E}\otimes\mathcal{H}_{\sf A}\ , \tag{261}$$

where $|\psi_{\rm in}\rangle \in \mathcal{H}_{\sf SA}$ is a purification of $\rho_{\rm in}$, U_θ a unitary on $\mathcal{H}_{\sf SE}$ and $|\epsilon_0\rangle$ a pure state on $\mathcal{H}_{\sf E}$ obtained from the Neumark extension theorem 3.3.1 (then $\{A_{\theta,i} = \langle\epsilon_i|U_\theta|\epsilon_0\rangle\}_{i=1}^p$ is a family of Kraus operators for $\mathcal{M}_\theta$). We set $|\Phi\rangle = |\Phi_{\theta_0}\rangle$ and $|\dot\Phi\rangle = {\rm d}|\Phi_\theta\rangle/{\rm d}\theta|_{\theta_0}$. Assuming that U_θ is $C^2(I)$, one has

$$\mathcal{F}_Q(|\Phi\rangle, \{|\Phi_\theta\rangle\}_{\theta\in I}) = 4\big\langle(\Delta H_{\theta_0})^2\otimes 1_{\sf A}\big\rangle_\Phi \quad , \quad H_\theta = {\rm i}\,\dot U_\theta U_\theta^\dagger \in \mathcal{B}(\mathcal{H}_{\sf SE})_{\rm s.a.}\ , \tag{262}$$

with $\Delta H_{\theta_0} = H_{\theta_0} - \langle H_{\theta_0}\otimes 1_{\sf A}\rangle_\Phi$. One can then show that

$$\begin{aligned}\mathcal{F}_Q(\rho_{\theta_0}, \{\mathcal{M}_\theta\}_{\theta\in I}) &= 4 \min_{B\in\mathcal{B}(\mathcal{H}_{\sf EA})_{\rm s.a.}} \Big\langle\big(\Delta H_{\theta_0}\otimes 1_{\sf A} - 1_{\sf S}\otimes\Delta B\big)^2\Big\rangle_\Phi &(263)\\ &= 4\,\Big\langle(\Delta H_{\theta_0})^2\otimes 1_{\sf A}\Big\rangle_\Phi - 4\,\Big\langle 1_{\sf S}\otimes(\Delta B^{\rm opt})^2\Big\rangle_\Phi\ , &(264)\end{aligned}$$

where $\rho_{\sf EA} = {\rm tr}_{\sf S}\,|\Phi\rangle\langle\Phi|$ is the reduced state of the environment and ancilla at the parameter value θ_0 and $\Delta B = B - \langle B\rangle_{\rho_{\sf EA}}$. The optimal self-adjoint operator $\Delta B^{\rm opt}$ on $\mathcal{H}_{\sf EA}$ minimizing the quantum expectation in (263) is given by the symmetric logarithmic derivative $(\mathcal{K}_{\sf B})_{\rho_{\sf EA}}^{-1}(D_{\sf EA})$ of the operator

$$D_{\sf EA} = \frac{\rm i}{2}\,{\rm tr}_{\sf S}\,\big(|\dot\Phi\rangle\langle\Phi| - |\Phi\rangle\langle\dot\Phi|\big) - \langle H_{\theta_0}\otimes 1_{\sf A}\rangle_\Phi\ \rho_{\sf EA}\ . \tag{265}$$

In other words, $\Delta B^{\rm opt}$ is a solution of the equation $\frac12\{\Delta B^{\rm opt}, \rho_{\sf EA}\} = D_{\sf EA}$. The first term in the right-hand side of (264) is the quantum Fisher information of the fixed family of purifications $\{|\Phi_\theta\rangle\}_{\theta\in I}$ on the enlarged Hilbert space; the second term can be interpreted as the extra amount of information that has leaked to the environment and must be substracted to obtain the quantum Fisher information of the system. In view of equation $(*)$ in the footnote after definition 7.6.1 and $|\dot\Phi\rangle = -{\rm i}H_{\theta_0}\otimes 1_{\sf A}|\Phi\rangle$, this term is equal to the quantum Fisher information $\mathcal{F}_Q(\rho_{\sf EA}, \{\sigma_\theta^{\sf EA}\}_{\theta\in I})$ for the following family of states on $\mathsf{H}_{\sf EA}$

$$\sigma_\theta^{\sf EA} = {\rm tr}_{\sf S}\,|\Sigma_\theta\rangle\langle\Sigma_\theta| \quad , \quad |\Sigma_\theta\rangle = \frac{(1+(\theta-\theta_0)H_{\theta_0}\otimes 1_{\sf A})|\Phi\rangle}{\|(1+(\theta-\theta_0)H_{\theta_0}\otimes 1_{\sf A})|\Phi\rangle\|}\ , \tag{266}$$

which can be prepared by performing a measurement on $\sf SE$ with Kraus operators $A_\theta^\pm = [1\pm(\theta-\theta_0)H_{\theta_0}]/\sqrt{2}$ conditioned to the outcome "+".

Let us sketch the proof of (263) and (264) (see [63] for more details). The first equation follows from a simple generalization of the proof of Proposition 7.6.3. The second equation follows from the first one and the fact that if ρ is a state and D a self-adjoint operator on $\mathcal{H}$ such that $P^\perp D P^\perp = 0$, $P^\perp$ being the projector onto $\ker(\rho)$, then the minimum of the function

$$f(B) = {\rm tr}(\rho B^2) - 2\,{\rm tr}(BD) \tag{267}$$

for all $B\in\mathcal{B}(\mathcal{H})_{\rm s.a.}$ such that $\langle B\rangle_\rho = 0$ is achieved for $B = B^{\rm opt} = (\mathcal{K}_{\sf B})_\rho^{-1}(D - {\rm tr}(D)\rho)$, and the minimum is given by $-\langle(B^{\rm opt})^2\rangle_\rho$.

Formula (263) can be used to bound $\mathcal{F}_Q(\rho_{\theta_0}, \{\mathcal{M}_\theta\}_{\theta\in I})$ from above by calculating the expectation value in the right-hand side for a suitable choice of the self-adjoint operator ΔB with zero mean. If ΔB is an approximate solution of $\frac12\{\Delta B, \rho_{\sf EA}\} = D_{\sf EA}$, this upper bound will approximate well the real value. This variational approach is of interest for instance when the quantum operations $\mathcal{M}_\theta$ on the system $\sf S$ originates from the coupling of $\sf S$ with its environment by a known θ-independent interaction

Hamiltonian H_{int}. Then one may take $U_\theta = e^{-\mathrm{i}\theta H_{\mathsf{S}}\otimes 1_{\mathsf{E}} - \mathrm{i}tH_{\text{int}} - \mathrm{i}t\, 1_{\mathsf{S}}\otimes H_{\mathsf{E}}}$ in (261), with t the interaction time and H_{E} the environment Hamiltonian. An application to the problem of phase estimation in the presence of phase diffusion is given in [63]. Note that, in contrast to formula (263), in this approach the diagonalization the density matrices ρ_{θ_0} is not required.

Remark 7.6.5. *A formula for $\mathcal{F}_Q$ involving a minimization over all families of pure state decompositions $\{|\psi_{i,\theta}\rangle, \eta_{i,\theta}\}_{i=1}^q$ of the output states ρ_θ, which generalizes (245) for general ρ_θ, is (see [69]):*

$$\mathcal{F}_Q(\rho_{\theta_0}, \{\mathcal{M}_\theta\}_{\theta\in I}) = 4 \min_{\{|\psi_{i,\theta}\rangle, \eta_{i,\theta}\}} \sum_{i=1}^q \left\| \dot{\tilde{\psi}}_{i,\theta_0} \right\|^2 = 4 \min_{\{|\psi_{i,\theta}\rangle, \eta_{i,\theta}\}} \left\{ \sum_{i=1}^q \left\| \dot{\tilde{\psi}}_{i,\theta_0} \right\|^2 - \left| \sum_{i=1}^q \langle \dot{\tilde{\psi}}_{i,\theta_0} | \tilde{\psi}_{i,\theta_0} \rangle \right|^2 \right\} \tag{268}$$

with $|\tilde{\psi}_{i,\theta}\rangle = \sqrt{\eta_{i,\theta}}|\psi_{i,\theta}\rangle$. It is not difficult to show that the right-hand side of this formula reduces to (263) by using the correspondance between pure state decompositions and purifications and the fact that two pure state decompositions are related to each other via a unitary matrix, see (14) and (16). The second equality in (268) follows from the observation that if $\{|\psi_{i,\theta_0}\rangle, \eta_{i,\theta_0}\}$ is a pure state decomposition of ρ_{θ_0} then $\{e^{-\mathrm{i}(\theta-\theta_0)\alpha}|\psi_{i,\theta_0}\rangle, \eta_{i,\theta_0}\}$ with $\alpha = -|\sum_i \langle \dot{\tilde{\psi}}_{i,\theta_0} | \widetilde{\psi}_{i,\theta_0}\rangle|^2$ is clearly also a pure state decomposition of ρ_{θ_0}.

7.6.3 Skew information

Definition 7.6.6. *Given any state $\rho \in \mathcal{E}(\mathcal{H})$ and observable $H \in \mathcal{B}(\mathcal{H})_{\text{s.a.}}$, the non-negative number*

$$\mathcal{I}_{\text{skew}}(\rho, H) = -\frac{1}{2} \operatorname{tr}\left([\sqrt{\rho}\,,\, H]^2 \right) \tag{269}$$

is called the skew information of ρ with respect to H.
If ρ is invertible then $\mathcal{I}_{\text{skew}}(\rho, H) = \frac{1}{2}(g_{\mathrm{H}})_\rho\big(-\mathrm{i}[H,\rho], -\mathrm{i}[H,\rho]\big)$.

The skew information was introduced by Wigner and Yanase [191]. It describes the amount of information on the values of observables not commuting with H that can be obtained from measurements on a system in state ρ. Like the quantum Fisher information, it is non-negative and vanishes if and only if $[\rho, H] = 0$. Furthermore, $\mathcal{I}_{\text{skew}}(\rho, H)$ is convex in ρ (this follows from the joint convexity of $d_{\mathrm{H}}(\rho,\sigma)^2$ by the same argument as in the proof of Proposition 7.6.3). The Fisher and skew informations are additive, in the sense that

$$\mathcal{F}_Q(\rho_1 \otimes \rho_2, H_1 \otimes 1 + 1 \otimes H_2) = \mathcal{F}_Q(\rho_1, H_1) + \mathcal{F}_Q(\rho_2, H_2)\ , \tag{270}$$

with a similar identity for $\mathcal{I}_{\text{skew}}$. This can be derived by using (239), (269), and the following identity on the matrix elements of the commutator $C = [\rho_1 \otimes \rho_2, H_1 \otimes 1 + 1 \otimes H_2]$ in the eigenbasis $\{|k_1\rangle|k_2\rangle\}$ of $\rho_1 \otimes \rho_2$:

$$\big|\langle k_1|\langle k_2|C|l_1\rangle|l_2\rangle\big|^2 = \delta_{k_2 l_2}\big(p_{k_2}^{(2)}\big)^2 \big|\langle k_1|[\rho_1, H_1]|l_1\rangle\big|^2 + \delta_{k_1 l_1}\big(p_{k_1}^{(1)}\big)^2 \big|\langle k_2|[\rho_2, H_2]|l_2\rangle\big|^2\ , \tag{271}$$

where $\{|k_1\rangle\}$ and $\{|k_2\rangle\}$ are eigenbases of ρ_1 and ρ_2 and $p_{k_1}^{(1)}$ and $p_{k_2}^{(2)}$ are the corresponding eigenvalues. One has the bounds

$$\frac{1}{8}\mathcal{F}_Q(\rho, H) \le \mathcal{I}_{\text{skew}}(\rho, H) \le \frac{1}{4}\mathcal{F}_Q(\rho, H) \le \langle (\Delta H)^2 \rangle_\rho\ , \tag{272}$$

where $\langle (\Delta H)^2 \rangle_\rho = \operatorname{tr}(\rho H^2) - (\operatorname{tr}\rho H)^2$ is the variance of H. The second and third inequalities are equalities for pure states. The two first inequalities in (272) follow from (224); the last one follows from the convexity of $\rho \mapsto \mathcal{F}_Q(\rho, H)$, (238), and the concavity of $\rho \mapsto \langle (\Delta H)^2 \rangle_\rho$.