Quantum Fisher Information: Variational principle and simple iterative algorithm for its efficient computation

Katarzyna Macieszczak
University of Nottingham, School of Mathematical Sciences, University Park, NG7 2RD Nottingham, UK

We derive a new variational principle for the quantum Fisher information leading to a simple iterative alternating algorithm, the convergence of which is proved. The case of a fixed measurement, i.e. the classical Fisher information, is also discussed.

I. INTRODUCTION

The Fisher information is an important concept in statistics. The Cramer-Rao inequality states that for any unbiased estimator $\widetilde{\varphi}$ of a parameter $\varphi \in \mathbb{R}$ its variance $\Delta^2\widetilde{\varphi}$ is lower bounded by the inverse of the Fisher information:

$$\Delta^2\widetilde{\varphi} = \int_X \mathrm{d}x\, p_\varphi(x)\,(\widetilde{\varphi}(x)-\varphi)^2 \geq F_\varphi^{-1}, \tag{1}$$

where the Fisher information F_φ is defined as follows:

$$F_\varphi = \int_{\{x\in X:\, p_\varphi(x)\neq 0\}} \mathrm{d}x\, p_\varphi(x) \left(\frac{\partial \log(p_\varphi(x))}{\partial\varphi}\right)^2. \tag{2}$$

Let x be a result of a POVM measurement $\{\Pi_x\}_{x\in X}$ ($\Pi_x \in \mathcal{B}(\mathcal{H})$, $\Pi_x \geq 0$, $\Pi_x = \Pi_x^\dagger$, $\int_X \mathrm{d}x\,\Pi_x = 1$) which is performed on a quantum state on a Hilbert space $\mathcal{H}$ described by a density matrix $\rho_\varphi \in S = \{\rho \in \mathcal{B}(\mathcal{H}) : \rho = \rho^\dagger, \mathrm{Tr}\{\rho\} = 1\}$, then $p_\varphi(x) = \mathrm{Tr}\{\rho_\varphi \Pi_x\}$. $F_{\rho_\varphi,\{\Pi_x\}}$ depends on the choice of measurement $\{\Pi_x\}_{x\in X}$, but whatever the measurement, we have:

$$F_{\rho_\varphi,\{\Pi_x\}} \leq F^Q_{\rho_\varphi} = \mathrm{Tr}\{\rho_\varphi L_\varphi^2\}, \quad \frac{1}{2}\{L_\varphi,\rho_\varphi\} = \frac{\mathrm{d}}{\mathrm{d}\varphi}\rho_\varphi, \tag{3}$$

where $F^Q_{\rho_\varphi}$ is the so called quantum Fisher information, L_φ is the symmetric logarythmic derivative and $\{.,.\}$ is the anticommutator. The eigenbasis of the L_φ operator corresponds to the optimal projective measurement for which the inequality in Eq. (3) is saturated.

Let us assume that $\rho_\varphi = e^{-i\varphi H}\rho\, e^{i\varphi H}$, where the generator H belongs to the $L^2(\mathcal{H})$ space of the self-adjoint Hilbert-Schmidt operators on $\mathcal{H}$. The quantum Fisher information has the same value for all φ, therefore let us drop the index φ. We have $F^Q_\rho = \mathrm{Tr}\{\rho L_\rho^2\}$ and $\frac{1}{2}\{L_\rho,\rho\} = -i[H,\rho]$.

II. VARIATIONAL PRINCIPLE

Let us consider the case in which the states used in the estimation are obtained as outputs from a quantum channel Λ (see Fig. 1), i.e these are elements of $\Lambda(S)$ where S is the set of density matrices. Unless Λ is a unitary channel,

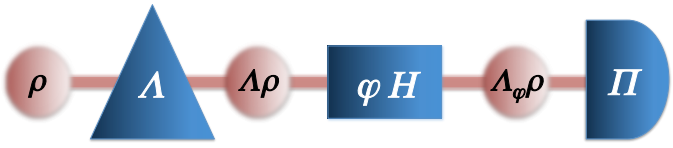

FIG. 1: The setup discussed in the draft: An input state ρ first goes through a quantum channel Λ, then a parameter value φ is imprinted using the H generator and finally a POVM measurement $\{\Pi_x\}_{x\in X}$ is performed to retrive the information about the parameter.

$\Lambda(S)$ is restricted and does not contain all density matrices, e.g. Λ can represent decoherence of a state ρ. In order to optimise the setup in Fig. 1 one is required to find $\sup_\rho F^Q_{\Lambda(\rho)}$ and the corresponding optimal input state ρ.

The supremum $F^Q = \sup_\rho F^Q_{\Lambda(\rho)}$ of the quantum Fisher information can be expressed by the following **variational principle**:

$$F^Q = \sup_{X\in L^2(\mathcal{H})} \sup_{|\psi\rangle} \langle\psi|\,\Lambda^\dagger\left(-X^2 + 2i[H,X]\right)|\psi\rangle, \tag{4}$$

where $|\psi\rangle$ is a normalised vector in $\mathcal{H}$. The set of normalised states can be replaced by the set S of all density matrices.

Proof. The order of optimisation is arbitrary. We have:

$$\sup_{X\in L^2(\mathcal{H})} \langle\psi|\,\Lambda^\dagger\left(-X^2+2i[H,X]\right)|\psi\rangle = \sup_{X\in L^2(\mathcal{H})} \mathrm{Tr}\{\Lambda\left(|\psi\rangle\langle\psi|\right)\,\left(-X^2+2i[H,X]\right)\}. \tag{5}$$

We perform differentation w.r.t. coefficients of X in any basis, thus arriving at the equation $\frac{1}{2}\{X,\Lambda\left(|\psi\rangle\langle\psi|\right)\} = -i[H,\Lambda\left(|\psi\rangle\langle\psi|\right)]$, which is just the equation of the symmetric logarithmic derivative $L_{\Lambda(|\psi\rangle\langle\psi|)}$. The matrix $F^{(2)}$ of second-order partial derivatives is negative, because when sandwiched between operators $Y\in L^2(\mathcal{H})$ it gives $Y^*F^{(2)}Y = -2\mathrm{Tr}\{\Lambda\left(|\psi\rangle\langle\psi|\right)Y^2\}$. Hence, this extremum is a maximum.

We are left with $\sup_{|\psi\rangle}\mathrm{Tr}\{\Lambda\left(|\psi\rangle\langle\psi|\right)L^2_{\Lambda(|\psi\rangle\langle\psi|)}\} = \sup_{|\psi\rangle} F^Q_{\Lambda(|\psi\rangle\langle\psi|)}$. The convexity of the quantum Fisher information w.r.t. density matrices implies that $\sup_{|\psi\rangle} F^Q_{\Lambda(|\psi\rangle\langle\psi|)} = \sup_\rho F^Q_{\Lambda(\rho)} = F^Q$. ■

III. ALGORITHM

Such a variational principle motivates the introduction of the following iterative alternating algorithm to effectively compute the maximum quantum Fisher information F^Q.

Algorithm. Let $F(\rho,X) := \mathrm{Tr}\{\rho\left(-X^2+2i[H,X]\right)\}$ and $G(X) := -X^2+2i[H,X]$ s.t. $F(\rho,X) = \mathrm{Tr}\{\rho\,G(X)\}$. One starts with the input state $|\psi_0\rangle$ and defines $\rho_0 = \Lambda(|\psi_0\rangle\langle\psi_0|)$. We know that the maximum value of $F(\rho_0,X)$ is attained for $X = L_{\rho_0}$. The input state $|\psi_1\rangle$ for the next step of the procedure is therefore chosen to be the eigenvector corresponding to the maximum eigenvalue of the operator $\Lambda^\dagger\left(G(L_{\rho_0})\right)$.

Analogously, let $|\psi_n\rangle$ be the initial state at n-th step and $\rho_n = \Lambda(|\psi_n\rangle\langle\psi_n|)$. We obtain $|\psi_{n+1}\rangle$ as the eigenvector corresponding to the maximum eigenvalue of the operator $\Lambda^\dagger\left(G(L_{\rho_n})\right)$. Therefore the quantum Fisher information of ρ_n increases with n:

$$F^Q_{\rho_n} = F(\rho_n,L_{\rho_n}) \le F(\rho_{n+1},L_{\rho_n}) \le F(\rho_{n+1},L_{\rho_{n+1}}) = F^Q_{\rho_{n+1}}. \tag{6}$$

We prove below that in the case of a finite dimension of $\mathcal{H}$ the algorithm provides the maximum quantum Fisher information $\lim_{n\to\infty} F^Q_{\rho_n} = F^Q$ and the optimal input state yield by a subsequence $\lim_{m\to\infty}|\psi_{n_m}\rangle$.

The maximization of the quantum Fisher information is thus achieved by alternatively 'moving' along two perpendicular 'directions' $L^2(\mathcal{H})$ and $\Lambda(S)$, where S is the set of density matrices. At each step we first go as high as possible in 'direction' $L^2(\mathcal{H})$ and then as high as possible in 'direction' $\Lambda(S)$. Since $F(\Lambda(\rho),X)$ is linear w.r.t. ρ, we always arrive at the boundary of S moving in the $\Lambda(S)$ direction, i.e. the chosen $\rho\in S$ is a pure state. The algorithm only requires diagonalising two operators: ρ_n and $\Lambda^\dagger\left(G(L_{\rho_n})\right)$ at each step.

Proof of convergence. The increasing sequence $f_n = F(\rho_n,L_{\rho_n}) = F^Q_{\rho_n}$ is bounded from above when $H\in L^2(\mathcal{H})$:

$$\sup_{|\psi\rangle} F^Q_{\Lambda(|\psi\rangle\langle\psi|)} \le \sup_\rho F^Q_\rho = \sup_{|\psi\rangle}\left(\langle\psi|H^2|\psi\rangle - \langle\psi|H|\psi\rangle^2\right) < \infty. \tag{7}$$

Thus, the $\{f_n\}$ sequence converges to a limit f^*.

First, let us prove that $F(\rho,X) = F(\rho,L_\rho) = F(\rho',L_{\rho'})$, where $\rho' = \Lambda(|\psi\rangle\langle\psi|)$ and $|\psi\rangle$ is the eigenvector corresponding to the maximum eigenvalue of $\Lambda^\dagger\left(G(L_\rho)\right)$, implies that $X = L_\rho$, $\rho = \rho'$ and ρ correponds to the maximum quantum Fisher information, i.e. $F^Q_\rho = F^Q$.

If L_ρ is uniquely defined for ρ, the equality $F(\rho,X) = F(\rho,L_\rho)$ implies $X = L_\rho$. On the other hand, the equation $F(\rho,L_\rho) = F(\rho',L_{\rho'})$ leads to $F(\rho,L_\rho) = F(\rho',L_\rho)$, which, if the maximal eigenvalue eigenspace of $\Lambda^\dagger\left(G(L_\rho)\right)$ is nondegenerate, imposes $\rho = \rho'$. We assume that, for the choice of L_ρ and $|\psi\rangle$ to be unique, these two conditions are fulfilled at all the steps of the algorithm (see Appendix B).

We prove now that ρ corresponds to the maximum, i.e. $F^Q_\rho = F^Q$:

$$\begin{aligned} F\left(\rho+\delta\rho^{(1)},L_\rho+\delta X^{(1)}\right) &= F\left(\rho+\delta\rho^{(1)},L_\rho\right) + \delta\,\mathrm{Tr}\left\{\rho\left(-\{L_\rho,X^{(1)}\}+2i[H,X^{(1)}]\right)\right\} + O(\delta^2) \\ &= F\left(\rho+\delta\rho^{(1)},L_\rho\right) + O(\delta^2) < F(\rho,L_\rho) + O(\delta^2), \end{aligned} \tag{8}$$

where the second equality is due to $\{L_\rho, \rho\} = 2i[H, \rho]$ and the cyclic property of the trace, and the inequality follows from ρ being the optimal state in $\Lambda(S)$ for $X = L_\rho$ ($\rho = \rho'$). Since $F(\rho, X)$ is concave w.r.t. (ρ, X), Eq. (8) implies that $F(\rho, L_\rho)$ is the global maximum F^Q.

If the algorithm gets stuck, i.e. $f_n = F(\rho_n, L_{\rho_n}) = F(\rho_{n+1}, L_{\rho_{n+1}}) = f_{n+1}$, we have arrived at the maximum F^Q, thus ending the proof for this case.

Let us consider the opposite case in which $f_n < f_{n+1}$ for all $n \in \mathbb{N}$. We assume that the dimension of H is finite from now on. As ρ_n are elements of S, which is compact, one can choose a convergent subsequence $\{\rho_{n_m}\}_{m\in\mathbb{N}}$; let ρ^* denote the limit of this subsequence. We assume that L_ρ is continuous w.r.t. ρ, then, together with the continuity of $F(\Lambda(\rho), X)$ w.r.t. (ρ, X), it implies $F(\rho^*, L_{\rho^*}) = \lim_{m\to\infty} F(\rho_{n_m}, L_{\rho_{n_m}}) = f^*$.

The eigenvector $|\psi^*\rangle$ corresponding to the maximum eigenvalue of $\Lambda^\dagger\left(G(L_{\rho^*})\right)$ leads to $\rho' = \Lambda(|\psi^*\rangle\langle\psi^*|)$ with the same quantum Fisher information $F^Q_{\rho'} = f^*$ as ρ' is the limit of the subsequence $\{\rho_{n_m+1}\}_{m\in\mathbb{N}}$. Therefore, if L_{ρ^*} is uniquely defined and $\Lambda^\dagger(G(L_{\rho^*})$ has a nondegenarate maximal eigenvalue eigenspace, we obtain $f^* = F(\rho^*, L_{\rho^*}) = F^Q$ according to the first part of the proof.

Therefore, we have proved the convergence to the maximum, i.e. $f^* = F^Q$, as long as, at each step, the choice of $|\psi_n\rangle$ and L_{ρ_n} is unique (see Appendix B). ■

The above proof is closely related to Chapter 10.3 in [1], where proof of convergence is presented for an alternating algorithm in the case of $\sup_{u_1\in A_1}\sup_{u_2\in A_2} f(u_1, u_2)$, where f is a strictly concave real-valued function and A_i is a compact and convex subset of $\mathbb{R}^{n_i}$, $i = 1, 2$. In our case the function $F(\Lambda(\rho), X)$ is linear w.r.t. ρ and strictly concave w.r.t. X, and, thanks to the definition of L_ρ, the $L^2(\mathcal{H})$ space in the variational principle in Eq. (4) can be restricted to the compact and convex set $B(0, ||H||_{HS})$ of operators with Hilbert-Schmidt norm not greater than $||H||_{HS}$ as $||L_\rho||_{HS} \leq ||H||_{HS}$.

IV. COMMENTS

A variational principle analogous to Eq. (4) holds for the maximum Fisher information $F_{\varphi,\{\Pi_x\}}$ for a fixed POVM measurement $\{\Pi_x\}_{x\in X}$:

$$F_{\varphi,\{\Pi_x\}} = \sup_{D\in L^2(X)} \sup_{|\psi\rangle} \langle\psi|\, \Lambda^\dagger_\varphi\left(-X_2^D + 2i[H, X_1^D]\right)|\psi\rangle, \tag{9}$$

where $X_j^D = \int_X \mathrm{d}x\, D(x)^j\, \Pi_x$, $\Lambda_\varphi(\rho) = e^{-i\varphi H}\Lambda(\rho)e^{i\varphi H}$, $L^2(X)$ is the set of square-integrable real-valued functions on X and $|\psi\rangle$ is a normalised vector in $\mathcal{H}$. If the measurement is a projective von Neumann measurement, $X_2 = X_1^2$.

The proof of Eq. (9) is presented in Appendix C. Here, let us simply state that for a given $|\psi\rangle$ the supremum $\sup_{D\in L^2(X)} \langle\psi|\, \Lambda^\dagger_\varphi\left(-X_2 + 2i[H, X_1]\right)|\psi\rangle$ is attained at:

$$D_L(x) = \begin{cases} \frac{\mathrm{Tr}\{-i[H, \Lambda_\varphi(|\psi\rangle\langle\psi|)]\Pi_x\}}{\mathrm{Tr}\{\Pi_x\Lambda_\varphi(|\psi\rangle\langle\psi|)\}}, & \text{if} \quad \mathrm{Tr}\{\Lambda_\varphi\left(|\psi\rangle\langle\psi|\right)\Pi_x\} > 0, \\ 0, & \text{elsewhere,} \end{cases} \tag{10}$$

and the supremum value equals $\langle\psi|\, \Lambda^\dagger_\varphi(X_2^{D_L})|\psi\rangle$.

Using Eq. (10) one can introduce an analogous alternating iterative algorithm to optimise the Fisher information for a fixed POVM measurement.

It can be easily demonstrated that the parabolic function $G(X) = \left(-X^2 + 2i[H, X]\right)$ appearing in Eq. (4) and in $F(\rho, X) = \mathrm{Tr}\{\rho\, G(X)\}$ can be replaced by any other function $G' : L^2(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ s.t. $\mathrm{Tr}\{\rho\, G'(X)\}$ has a global maximum w.r.t X equal to $\mathrm{Tr}\{\rho\, X_\rho^2\}$, where $\frac{1}{2}\{X_\rho, \rho\} = -i[H, \rho]$.

One can consider a **general case** when the parameter φ is encoded on a state ρ via a channel Λ_φ, i.e. $\rho_\varphi = \Lambda_\varphi(\rho)$. This decribes e.g. decoherence channels which do not commute with H: $\Lambda_\varphi(\rho) \neq e^{-i\varphi H}\Lambda_0(\rho)e^{i\varphi H}$. Let Λ'_φ denote $\frac{\mathrm{d}}{\mathrm{d}\varphi}\Lambda_\varphi$. We obtain the following variational principle:

$$F^Q = \sup_{X\in L^2(\mathcal{H})} \sup_{|\psi\rangle} \langle\psi|\, \left(-\Lambda^\dagger_\varphi(X^2) + 2\Lambda'^\dagger_\varphi(X)\right)|\psi\rangle, \tag{11}$$

where $|\psi\rangle$ is a normalised vector in $\mathcal{H}$. The set of normalised states can be replaced by the set of density matrices.

The general variational principle in Eq. (11) suggests introducing an alternating iterative algorithm analogous to the one discussed above. In the case of a fixed POVM measurement Eq. (11) can also easily be modified.

If we restrict the set of normalised states, or equally, the set of density matrices in Eq. (4) to a **class of states**, e.g. MPS states or gaussian states, we obtain a variational principle for the maximum quantum Fisher information w.r.t. this class C:

$$\sup_{D\in L^2(X)}\sup_{\rho\in C} F(\rho,X) = \sup_{\rho\in C}\sup_{X\in L^2(\mathcal{H})} F(\rho,X) = \sup_{\rho\in C} F_\rho^Q. \tag{12}$$

Furthermore, if C is convex, an alternating iterative algorithm can be used. All this is also true for the variational principles in Eq. (9) and (11).

Initially, we obtained the variational principle in Eq. (4) using Bayesian estimation in the limit of a deterministic prior distribution, which is explained in Appendix A.

ACKNOWLEDGEMENTS

The author is grateful to Rafał Demkowicz-Dobrzański for fruiftul discussions on Bayesian estimation and Marcin Jarzyna for checking the efficient performance of the algorithm. Helpful comments on this draft by Mădălin Guţă and Luis A. Correa are gratefully acknowledged. This research was supported by Polish NCBiR under the ERA-NET CHIST-ERA project QUASAR.

[1] R. W. Yeung *A First Course in Information Theory (Information Technology: Transmission, Processing and Storage)* (Springer-Verlag New York, 2006).
[2] K. Macieszczak, R. Demkowicz-Dobrzański, M. Fraas, Optimal quantum frequency estimation, *ArXiv e-Prints* (2013), arXiv:1311.5576 [quant-ph].
[3] R. Demkowicz-Dobrzanski, Beyond quantum Fisher information: optimal phase estimation with arbitrary a priori knowledge, *Phys. Rev. A* **83**, 061802 (2011).

Appendix A: Fisher information in Bayesian estimation

For the Bayesian estimation with a prior distribution $g(\varphi)$, we have an interesting connection with the Fisher information in the case of the Gaussian g. Let us assume, without loss of generality, that the mean of this distribution equals 0. Let Δ^2_{prior} denote the prior variance $\int \mathrm{d}\varphi\, g(\varphi)\,\varphi^2$.

The average variance of an estimator $\widetilde{\varphi}$ is defined as:

$$\Delta^2\widetilde{\varphi} = \int \mathrm{d}\varphi\, g(\varphi) \int \mathrm{d}x\, p_\varphi(x)\, \left(\widetilde{\varphi}(x)-\varphi\right)^2. \tag{A1}$$

In the case of Bayesian estimation the best estimator is known to be the conditional expected value, defined as follows:

$$\widetilde{\varphi}(x) = \begin{cases} \frac{\int \mathrm{d}\varphi\, g(\varphi)\, p_\varphi(x)\,\varphi}{\int \mathrm{d}\varphi\, g(\varphi)\, p_\varphi(x)}, & \text{if} \quad \int \mathrm{d}\varphi\, g(\varphi)\, p_\varphi(x) > 0, \\ 0, & \text{elsewhere.} \end{cases} \tag{A2}$$

We have:

$$\Delta^2\widetilde{\varphi} = \Delta^2_{prior} - \int \mathrm{d}\varphi\, g(\varphi) \int \mathrm{d}x\, p_\varphi(x)\, \widetilde{\varphi}(x)^2. \tag{A3}$$

For the gaussian distribution $g(\varphi) \propto \exp(-\frac{\varphi^2}{2\Delta^2_{prior}})$ we have $g(\varphi)\,\varphi = \Delta^{-2}_{prior}\frac{\mathrm{d}\,g(\varphi)}{\mathrm{d}\varphi}$. Thus:

$$\Delta^{-2}_{prior}\,\widetilde{\varphi}(x) = \frac{\int \mathrm{d}\varphi \frac{\mathrm{d}\,g(\varphi)}{\mathrm{d}\varphi}\,p_\varphi(x)}{\int \mathrm{d}\varphi\, g(\varphi)\,p_\varphi(x)} = -\frac{\int \mathrm{d}\varphi\, g(\varphi)\frac{\mathrm{d}\,p_\varphi(x)}{\mathrm{d}\varphi}}{\int \mathrm{d}\varphi\, g(\varphi) p_\varphi(x)} \quad \text{and} \tag{A4}$$

$$\begin{aligned}\frac{1-\frac{\Delta^2\widetilde{\varphi}}{\Delta^2_{prior}}}{\Delta^2_{prior}} &= \frac{\int \mathrm{d}\varphi\, g(\varphi)\int \mathrm{d}x\, p_\varphi(x)\,\widetilde{\varphi}(x)^2}{\Delta^4_{prior}} \\ &= \int \mathrm{d}x \int \mathrm{d}\varphi\, g(\varphi) p_\varphi(x) \left(\frac{\int \mathrm{d}\varphi\, g(\varphi)\frac{\mathrm{d}\,p_\varphi(x)}{\mathrm{d}\varphi}}{\int \mathrm{d}\varphi\, g(\varphi) p_\varphi(x)}\right)^2 = F^g_0,\end{aligned} \tag{A5}$$

where F^g_0 is the Fisher information at $\phi = 0$ for a new family of distributions $p^g_\phi(x) = \int \mathrm{d}\varphi\, g(\varphi) p_{\varphi+\phi}(x)$. If we choose from different families of distributions $\{p_\varphi\}$, e.g. various initial states ρ, minimisation of $\Delta^2\widetilde{\varphi}$ is equivalent to maximisation of F^g_0.

In the limit $g(\varphi) \to \delta(\varphi)$:

$$\lim_{\Delta_{prior}\to 0} F^g_0 = \int \mathrm{d}x\, p_0(x) \left(\frac{\frac{\mathrm{d}\,p_\varphi(x)}{\mathrm{d}\varphi}|_{\varphi=0}}{p_0(x)}\right)^2 = F_0. \tag{A6}$$

For the setup described in the main text we have $\rho_\varphi = e^{-i\varphi H}\rho e^{i\varphi H}$ and $p_\varphi(x) = \mathrm{Tr}\{\Pi_x \rho_\varphi\}$, where $\rho = \Lambda(|\psi\rangle\langle\psi|)$ is the initial state and $\{\Pi_x\}_{x\in X}$ describes the measurement. For a prior distribution g we look for a choice of ρ, $\{\Pi_x\}_{x\in X}$ and $\widetilde{\varphi}$ which minimises $\Delta^2\widetilde{\varphi}$. The case of φ being the frequency difference in an atomic clock was discussed in [2], where the relation in Eq. (A5) above is already reffered to in Eq. (6). In the case of a fixed measurement, let F^g_ρ denote F^g_0 for the initial state ρ.

To optimise the setup with a fixed measurement, we perform:

$$\sup_{|\psi\rangle} F^g_{\Lambda(|\psi\rangle\langle\psi|)} = \sup_{|\psi\rangle}\sup_{\widetilde{\varphi}} \frac{1-\frac{\Delta^2\widetilde{\varphi}}{\Delta^2_{prior}}}{\Delta^2_{prior}}. \tag{A7}$$

By exchanging the suprema and taking the limit $\Delta_{prior} \to 0$ we arrive at the variational principle for $F_{\{\Pi_x\}}$ in Eq. (9). A function D^L is the limit of best Bayesian estimators for $\rho = \Lambda(|\psi\rangle\langle\psi|)$.

When the measurement is also being optimised, we arrive at Eq. (4) and the symmetric logarithmic derivative of $\Lambda(|\psi\rangle\langle\psi|)$ is the limit of the operators $\int_X \mathrm{d}x\,\widetilde{\varphi}(x)\,\Pi_x$ encoding the best Bayesian estimators and projective measurements for this state.

In [3] an alternating iterative algorithm was used to optimise the initial states and measurements in phase estimation on the $[-\pi, \pi]$ interval with a given prior distribution and a sinusoidal cost function. In [2] an analogous algorithm was introduced for frequency estimation with a given prior distribution and a square cost function. In this draft an alternating iterative algorithm, to optimise the Fisher information, in both cases of a measurement being fixed or capable of being optimised, is both proposed and proved.

Appendix B: Unique choices in the algorithm

As stated above, the unique choices of $|\psi_n\rangle$ and L_{ρ_n} at each step are cruicial.

Let us note that L_ρ is uniquely defined only when ρ has a maximum possible rank. Let $\rho = \sum_{j=1}^{d'} \lambda_j\,|\lambda_j\rangle\langle\lambda_j|$, where d' is the rank of ρ and $\{|\lambda_j\rangle\}_{j=1}^{d'}$ are orthonormal vectors in $\mathcal{H}$. It can be shown that the equation $\frac{1}{2}\{X_\rho, \rho\} = -i[H, \rho]$ imposes the following form of L_ρ:

$$L_\rho = \sum_{j,k=1}^{d'} \left(\frac{\lambda_k - \lambda_j}{\lambda_k + \lambda_j}\langle\lambda_k|iH|\lambda_j\rangle\right) - P_{V^\perp} iH P_V + P_V iH P_{V^\perp} + L^a_\rho\,. \tag{B1}$$

where $V^\perp$ is the orthogonal complement of V, P_V, $P_{V^\perp}$ are the orthogonal projections on V and $V^\perp$ respectively and $L^a_\rho : V^\perp \to V^\perp$ represents the part of L_ρ which can be defined arbitrarily. From Eq. (B1) we see that L_ρ is only uniquely defined when $V = \mathcal{H}$.

If L_ρ is not uniquely defined, we set $L^a_\rho = 0$ in the algorithm. In the case of a unitary channel Λ, the state $\Lambda(|\psi\rangle\langle\psi|)$ has rank equal to 1. Nevertheless, we observed numerically that as long as the initial state $|\psi_0\rangle$ has non-zero expansion coefficients in the eigenbasis of H, the algorithm converges to the maximum value $F^Q = \sup_{|\psi\rangle}\left(\langle\psi|H^2|\psi\rangle - \langle\psi|H|\psi\rangle^2\right)$. This condition is related to the fact that when ρ_n has a block diagonal form in the eigenbasis of H, L_{ρ_n} also has an analogous block diagonal form according to Eq. (B1) and, moreover, this form is preserved at all the following steps of the algorithm. Assuming the nondegeneracy of the maximal eigenvalue of $\Lambda^\dagger(L_{\rho_n})$, ρ_{n+1} is only supported on one of the blocks, $L_{\rho_{n+1}}$ is not uniquely defined and, from this point on, the algorithm is effectively restricted to the subspace of $\mathcal{H}$ related to this block. Therefore, in such a case the algorithm will fail to provide the maximum quantum Fisher information F^Q, unless that maximum corresponds to the state which is supported only on this block.

In the case of a nonunitary Λ which commutes with H, i.e. $\Lambda(e^{-i\varphi H}\rho e^{i\varphi H}) = e^{-i\varphi H}\Lambda(\rho)e^{i\varphi H}$, a block diagonal form of ρ_n in the eigenbasis of H is also preserved by the algorithm. Therefore, in both these cases we make the following **proposition**.

If at each step of the algorithm ρ_n is irreducible w.r.t. the direct sums of the eigenspaces of H, the algorithm will converge to the maximum value F^Q.

By saying that ρ is irreducible w.r.t. the direct sums of the eigenspaces of H, we understand the following: if for a subspace $V \subset \mathcal{H}$ we have both $\rho V \subset V$ and $HV \subset V$, then either $V = \{0\}$ or $V = \mathcal{H}$. This is equivalent to saying that ρ has a block diagonal form in the eigenbasis of H which consists just of one block.

Appendix C: Proof of the variational principle in the case of a fixed measurement

Let $F_{\Lambda_\varphi(|\psi\rangle\langle\psi|),\{\Pi_x\}}$ be the Fisher information for a state $\Lambda_\varphi(|\psi\rangle\langle\psi|)$ and a measurement $\{\Pi_x\}_x \in X$. It is enough to prove that for a normalised vector $|\psi\rangle$ in $\mathcal{H}$:

$$\sup_{D\in L^2(X)} \langle\psi|\Lambda^\dagger_\varphi\left(-X_2^D + 2i[H, X_1^D]\right)|\psi\rangle = \langle\psi|\Lambda^\dagger_\varphi\left(-X_2^{D_L} + 2i[H, X_1^{D_L}]\right)|\psi\rangle, \tag{C1}$$

where D_L is defined in Eq. (10), and that

$$F_{\Lambda_\varphi(|\psi\rangle\langle\psi|),\{\Pi_x\}} = \langle\psi|\Lambda^\dagger_\varphi\left(-X_2^{D_L} + 2i[H, X_1^{D_L}]\right)|\psi\rangle. \tag{C2}$$

Once more, the convexity of the Fisher information w.r.t. a density matrix, guarantees $F_{\varphi,\{\Pi_x\}} = \sup_{|\psi\rangle} F_{\{\Pi_x\},\Lambda_\varphi(|\psi\rangle\langle\psi|)}$.

The equality in Eq. (C1) is proved by taking either the derivative w.r.t. $D(x)$ when X is discrete, or a functional derivative w.r.t. to $D(x)$ in a continuous case. $D_L(x)$ can be defined arbitrarily when $\mathrm{Tr}\{\Pi_x\Lambda_\varphi\left(|\psi\rangle\langle\psi|\right)\} = 0$.

In Eq. (C2), the identity is a consequence of the definition of the D_L function:

$$\begin{aligned}
\langle\psi|\Lambda^\dagger_\varphi\left(-X_2^{D_L} + 2i[H, X_1^{D_L}]\right)|\psi\rangle &= \mathrm{Tr}\{\Lambda_\varphi(|\psi\rangle\langle\psi|)X_2^{D_L}\} \\
&= \int_{X_{\Lambda_\varphi}} \mathrm{d}x\, \mathrm{Tr}\{\Pi_x\Lambda_\varphi\left(|\psi\rangle\langle\psi|\right)\}\left(\frac{\mathrm{Tr}\{-i[H, \Lambda_\varphi\left(|\psi\rangle\langle\psi|\right)]\Pi_x\}}{\mathrm{Tr}\{\Lambda_\varphi\left(|\psi\rangle\langle\psi|\right)\Pi_x\}}\right)^2 \\
&= \int_{\{x:\, p_\varphi(x)\neq 0\}} \mathrm{d}x\, p_\varphi(x)\left(\frac{\frac{\partial p_\varphi(x)}{\partial\varphi}}{p_\varphi(x)}\right)^2 = F_{\Lambda_\varphi(|\psi\rangle\langle\psi|),\{\Pi_x\}}
\end{aligned} \tag{C3}$$

where $X_{\Lambda_\varphi} = \{x :\, \mathrm{Tr}\{\Lambda_\varphi\left(|\psi\rangle\langle\psi|\right)\Pi_x\} \neq 0\}$ and $p_\varphi(x) = \mathrm{Tr}\{\Lambda_\varphi\left(|\psi\rangle\langle\psi|\right)\Pi_x\}$. ■