

Quantum Measurement and Estimation Theory

Rafał Demkowicz-Dobrzański

April 23, 2021

Contents

I	Quantum preliminaries	7
1	States, evolution, measurement	9
2	Generalized measurements	11
2.1	Stern-Gerlach experiment	11
2.2	Mathematical formulation	12
2.3	Post-measurement state	14
2.4	Decoherence and completely positive maps	14
3	Uncertainty relations	17
3.1	Heisenberg-Robertson uncertainty relation	17
3.2	Precision vs. Disturbance uncertainty relation	17
3.3	Joint measurement	18
II	Classical estimation theory	23
4	Frequentist approach	27
4.1	Optimal unbiased estimator	27
4.2	Cramér-Rao bound	28
4.3	Multi-parameter case	29
4.4	Maximum likelihood estimator	30
5	Bayesian approach	35
5.1	Optimal Bayesian estimator	35
5.2	Bayesian Cramér-Rao bound	36
III	Quantum estimation theory	39
6	Quantum frequentist estimation	43
6.1	Quantum Cramér-Rao bound	43
6.2	Multi-parameter case	46
6.3	Natural metric in the space of quantum states	47
6.4	Holevo Cramer-Rao bound	49

7	Quantum Bayesian estimation	53
7.1	Quadratic cost problems	54
7.2	Covariant estimation problems	55
IV	Quantum metrology	63
8	Noiseless quantum metrology	67
8.1	Frequentist approach	67
8.1.1	N parallel channels	68
8.1.2	General adaptive strategy	68
8.2	Bayesian approach	68
8.2.1	General adaptive strategy	70
8.3	Phase vs. Frequency estimation	70
9	Impact of decoherence	73
9.1	Parallel channels	74
9.2	General adaptive protocols	76
9.3	Constrained protocols	76
10	Practical protocols	79
11	General Markovian noise models	81
12	Quantum error-correction	83
13	Waveform estimation	85
14	Adiabatic quantum metrology	87
V	And beyond	89
15	Quantum discrimination	91
16	Quantum communication	93
17	Quantum computing	95
18	Quantum thermodynamics	97

Genesis

Literature:

- C. W. Helstrom, *Quantum Detection and Estimation Theory*, (1976)
- A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, (1982)
- S. M. Kay, *Fundamentals of statistical signal processing* (1993)
- R. Demkowicz-Dobrzanski, M. Jarzyna, J. Kolodynski, *Quantum limits in optical interferometry*, Progress in Optics 60, 345 (2015)

Part I

Quantum preliminaries

Chapter 1

States, evolution, measurement

Pure states, density matrices, reduced density matrix, unitary evolution, projective measurement

Chapter 2

Generalized measurements

Figure 2.1: Stern-Gerlach experiment

2.1 Stern-Gerlach experiment

In order to understand the need of introducing the concept of generalized measurement, let us start with a detailed analysis of a model of the Stern-Gerlach experiment, see Fig. 2.2. Consider a spin $\frac{1}{2}$ particle, initial state of which is given by:

$$|\psi\rangle = |s\rangle \otimes |\phi\rangle_z, \quad (2.1)$$

where $|s\rangle$ represents a general spin state of the particle, while $|\phi\rangle_z$ is its spatial wave function, which for the purpose of this model we restrict to represent solely the z coordinate degree of freedom. The particle travels in the x direction (the movement in the x direction we treat classically) and enters a nonuniform magnetic field at the origin of the coordinate frame, which can be approximated as ¹:

$$\vec{B} \approx (B_0 + kz)\hat{e}_z. \quad (2.2)$$

The Hamiltonian that describes the interaction of the particle with the magnetic field is given by

$$H = -\mu\vec{\sigma} \cdot \vec{B}, \quad (2.3)$$

¹This is just an approximation, since according to Maxwell's equations magnetic field has to satisfy $\vec{\nabla} \cdot \vec{B} = 0$, while these field clearly does not satisfy this condition. In reality we would need to take into account other magnetic field components to satisfy the divergence-free requirement

where $\vec{\sigma} = [\sigma_x, \sigma_y, \sigma_z]^T$ is a vector of Pauli matrices. We assume that the interaction lasts for a short time δt during which the particle goes through the magnetic field. After this time the interaction is not present anymore. In order to further simplify the formulas we will set $B_0 = 0$. Before the interaction takes place, the initial state of the particle corresponds to

$$|s\rangle = c_+|+\frac{1}{2}\rangle_z + c_-|-\frac{1}{2}\rangle_z, \quad (2.4)$$

$$|\phi\rangle_z = \frac{1}{\sqrt[4]{2\pi\sigma^2}} \int dz e^{-\frac{z^2}{4\sigma^2}} |z\rangle, \quad (2.5)$$

where the general spin state is written in terms of states with a definite spin projection on the z axis, while the initial spatial wave function is assumed to be gaussian with a mean deviation σ . In what follows we will for simplicity denote $|\pm\rangle := |\pm\frac{1}{2}\rangle_z$. Let us evolve the state for time δt under the action of Hamiltonian H (we ignore particle free evolution, on the grounds that δt is sufficiently small):

$$|\psi(\delta t)\rangle = e^{-\frac{iH\delta t}{\hbar}} |\psi\rangle = c_+|+\rangle|\phi_+\rangle + c_-|-\rangle|\phi_-\rangle, \quad (2.6)$$

where

$$|\phi_\pm\rangle = \frac{1}{\sqrt[4]{2\pi\sigma^2}} \int dz e^{-\frac{z^2}{4\sigma^2} \pm \frac{i\mu k \delta t}{\hbar} z} |z\rangle. \quad (2.7)$$

In order to interpret the above states, let us write them in the momentum representation:

$$|\phi_{\pm}\rangle = \frac{1}{\sqrt{4\pi\sigma_p^2}} \int dp e^{-\frac{(p \mp \delta p)^2}{4\sigma_p^2}} |p\rangle, \quad (2.8)$$

where $\sigma_p = \frac{\hbar}{2\sigma}$ is the width of the gaussian wave packet in the momentum representation, while $\delta p = \mu k \delta t$ represents the momentum kick experienced by the particle.

We see that as a result of the evolution, the spin and spatial degrees of freedom become entangled and the particle experiences a momentum kick that depends on the projection of its spin on the z axis. If we now measure the momentum of the particle we will learn some information on the spin state

Let us assume we have performed an ideal projective measurement of the momentum of the particle². The probability of obtaining result p can be calculated using the following formula (watch out for a slight abuse of notation...):

$$p(p) = \langle \psi(\delta t) | \mathbb{1} \otimes |p\rangle \langle p| | \psi(\delta t) \rangle, \quad (2.9)$$

where the identity reminds us that we do not measure the spin states of the particle directly. Explicitly it reads:

$$p(p) = \frac{1}{\sqrt{2\pi\sigma^2}} \left(|c_+|^2 e^{-\frac{(p-\delta p)^2}{2\sigma_p^2}} + |c_-|^2 e^{-\frac{(p+\delta p)^2}{2\sigma_p^2}} \right) \quad (2.10)$$

We see that the probability distribution is given solely in terms of the spin degrees of freedom parameters and the measurement result p .

The idea of the generalized measurement formalism is to forget the details of the whole interaction between the measured system (here the spin degree of freedom) and the measuring device (here the spatial degree of

Figure 2.2: Conceptual scheme of a generalized measurement

freedom), and express the resulting probability distribution in the form:

$$p(p) = \langle s | \Pi_p | s \rangle, \quad (2.11)$$

where Π_p are the respective measurement operators, which need not in general be projective operators. Still they need to be positive (in order for the probability distribution to be positive) and sum up to identity $\int dp \Pi_p = \mathbb{1}$. In our case the corresponding operators are easy to find and read explicitly:

$$\Pi_p = \frac{1}{\sqrt{2\pi\sigma^2}} \begin{bmatrix} e^{-\frac{(p-\delta p)^2}{2\sigma_p^2}} & 0 \\ 0 & e^{-\frac{(p+\delta p)^2}{2\sigma_p^2}} \end{bmatrix} \quad (2.12)$$

when written in the $\{|+\rangle, |-\rangle\}$ basis.

Note that in case $\delta p \ll \sigma_p$ the measurement provide us with almost no information, while in the opposite case the measurement results are highly informative regarding the spin information

2.2 Mathematical formulation

Inspired by the Stern-Gerlach example, we are now ready to present the general formulation of the concept of generalized measurements.

Consider two quantum systems: the system to be measured (S) and the measuring device (M). The general idea of a generalized measurement, is to let the system interact with the measuring device, after which the measurement device state is read out using a standard projective measurement. Initially, the system and the measurement device are uncorrelated and their state is given by, $\rho_{SM} = \rho_S \otimes |0\rangle\langle 0|_M$, where we assumed (without loss of generality, as we may always purify the measuring device system to a larger space) that the M is prepared initially in a pure state. As a result of a unitary interaction of the two

²In a real Stern-Gerlach experiment, we let the particle evolve for some time t , and then measure the position when the particle hits the screen. If the time t is long enough this will be equivalent to the measurement of the momentum of the particle—see Problem 2.1

systems the final state reads:

$$\rho'_{SM} = U \rho_{SM} U^\dagger. \quad (2.13)$$

Finally, a measurement, projecting the state of M onto a $\{|i\rangle_M\}$ basis ($P_i = |i\rangle\langle i|_M$) is performed resulting in the probability distribution:

$$\begin{aligned} p(i) &= \text{Tr}_{SM} (\rho'_{SM} \mathbb{1}_S \otimes P_i) = \\ &= \text{Tr}_{SM} \left(U \rho_S \otimes |0\rangle\langle 0|_M U^\dagger \mathbb{1} \otimes |i\rangle\langle i| \right) = \\ &= \text{Tr}_S (\rho_S \Pi_i), \end{aligned} \quad (2.14)$$

where $\Pi_i = {}_M \langle 0|U^\dagger \mathbb{1} \otimes P_i U|0\rangle_M$. Π_i are generalized measurement operators, which in general need not be projective operators. Still from the above construction it follows that they satisfy positivity ($\Pi_i \geq 0$) and completeness ($\sum_i \Pi_i = \mathbb{1}$) conditions.

Any set of operators, $\{\Pi_i\}$ such that $\Pi_i \geq 0$, $\sum_i \Pi_i = \mathbb{1}$ is called a positive operator valued measure (POVM), since we may regard it as an operator which when traced with the density matrix generates a measure in the space of events i . We have seen above, that the interaction of a quantum system with a measuring system followed by the projective measurement of the latter results in an effective description of the measurement using a POVM.

The question, which is essential for the whole field of quantum estimation theory, is whether for any given POVM there is a physical realization, in the sense of a particular form of interaction between S and M and a particular measurement that results in the effective probability distribution described by this POVM. The answer is yes, and it is known under the name of Naimark dilation theorem. We provide the finite-dimensional case proof below.

Let $\{\Pi_i\}$, $i = 1, \dots, K$ be a POVM, $\Pi_i \in \mathcal{L}(\mathbb{C}_d)$. We will show, that there exist a unitary $U \in \mathcal{L}(\mathbb{C}_{dK})$ and a projective measure-

ment P_i on \mathbb{C}_K such that,

$$\forall_{\rho_S} \text{Tr}(\rho_S \Pi_i) = \text{Tr}(U \rho_S \otimes |0\rangle\langle 0| U^\dagger \mathbb{1} \otimes P_i). \quad (2.15)$$

Let us define

$$U|\psi\rangle \otimes |0\rangle = \sum_{i=1}^K \sqrt{\Pi_i} |\psi\rangle \otimes |i\rangle. \quad (2.16)$$

Note, that there is no problem in taking a square root from Π_i as it is a positive operator. Note also, that if this U is a legitimate unitary operation, then when accompanied by the projection $P_i = |i\rangle\langle i|$, realizes the required POVM. Consider the above operation U acting on two different input states. U preserves scalar products between the states, as

$$\begin{aligned} \left(\sum_i \langle \psi' | \otimes \langle i | \sqrt{\Pi_i} \right) \left(\sum_j \sqrt{\Pi_j} |\psi'\rangle \otimes |j\rangle \right) &= \\ \langle \psi' | \sum_i \Pi_i |\psi\rangle &= \langle \psi' | \psi \rangle. \end{aligned} \quad (2.17)$$

The above property is not in itself a sufficient condition for unitarity (it is necessary), since we only defined the map on a subset of state in the whole Hilbert space \mathbb{C}_{dK} (the subset of states of the form $|\psi\rangle \otimes |0\rangle$). In other words, we have shown, that if we write U as a matrix, then the d columns are legitimate columns take from a unitary matrix—they are orthonormal to each other. As such, we may always perform a completion of the matrix to a full unitary matrix by adding additional orthogonal vectors until we get all the columns which constitute the the whole orthonormal basis. This ends the proof ■.

Is U in the above construction unique? No. We could as well take:

$$U|\psi\rangle \otimes |0\rangle = \sum_{i=1}^K V_i \sqrt{\Pi_i} |\psi\rangle \otimes |i\rangle, \quad (2.18)$$

where V_i are unitaries. This unitaries, may be understood as rotations of the post-measurement state—they do not affect the

probabilities of obtaining different measurement results.

2.3 Post-measurement state

If instead of just calculating the probabilities, we wanted to write down the post-measurement state itself, we need to go back to the representation of the generalized measurement as a subsequent interaction with the measuring device and a projective measurement on it. Given measurement result i the joined output state of the S and M subsystems reads:

$$\begin{aligned} \rho_{SM}'^{(i)} &= \mathbb{1} \otimes P_i \rho_{SM}' \mathbb{1} \otimes P_i = \\ & \mathbb{1} \otimes |i\rangle\langle i| U \rho \otimes |0\rangle\langle 0| U^\dagger \mathbb{1} \otimes |i\rangle\langle i|. \end{aligned} \quad (2.19)$$

Tracing out the M subsystem we get the conditional state

$$\rho_S'^{(i)} = \langle i|U|0\rangle \rho_S \langle 0|U^\dagger|i\rangle = K_i \rho_S K_i^\dagger, \quad (2.20)$$

where we have introduced the so called Kraus operators $K_i = \langle i|U|0\rangle$ (note that this is a partial scalar product, that leaves an operator acting on the S system). This state is subnormalized and its trace gives the probability of obtaining result i :

$$\begin{aligned} p(i) &= \text{Tr}(\rho_S'^{(i)}) = \text{Tr}(K_i \rho_S K_i^\dagger) = \\ &= \text{Tr}(\rho_S K_i^\dagger K_i) = \text{Tr}(\rho_S \Pi_i) = p(i), \end{aligned} \quad (2.21)$$

where we have used the property that $\Pi_i = K_i^\dagger K_i$. If we insist on writing a normalized conditional state, we should write $\rho_S'^{(i)}/p_i$.

Note, that given Π_i the corresponding K_i are determined only up to a unitary: $K_i = V_i \sqrt{\Pi_i}$, where V_i can be arbitrary unitary. This represents that fact, that after the measurement result is obtained, one may freely rotate the state depending on the measurement results, and this freedom does not appear in the statistics of the measurement results.

Recalling the example of the Stern-Gerlach experiment, one can see that in the limit

$\delta p \ll \sigma_p$ the Π_p operators are practically proportional to identity. This implies that the while we get very little information about the spin, the spin state is also almost not disturbed at all. We refer to such a regime as the *weak measurement* regime. In the opposite case where we obtain a lot of information but at the same time disturb the state we say we deal with a *strong measurement*.

2.4 Decoherence and completely positive maps

Imagine now a situation in which, in the above described protocol, we forget to register the actual measurement result. We can regard this situation in a spirit, that we simply do not have access to the readout of the measurement performed on the subsystem M , or in other words that M should be treated as inaccessible environment with which our system S interacts. In such a situation, the output state of the system S is obtained by simply tracing out the joined state of S and M over subsystem M , and reads:

$$\rho_S' = \text{Tr}_M \left(U \rho_S \otimes |0\rangle\langle 0| U^\dagger \right) = \sum_i K_i \rho_S K_i^\dagger. \quad (2.22)$$

The above formula has a clear intuitional meaning. This is a mixture of different conditional states corresponding to different measurement results i , representing the fact that we have no knowledge of the actual value of i and hence we are forced to consider the mixture only.

The above formula represents a general structure of a quantum channel. Kraus operators, K_i , can be arbitrary operators (not necessary unitary, hermitian, ...), but in order to guarantee the trace-preservation condition they need to satisfy: $\sum_i K_i^\dagger K_i = \mathbb{1}$. Note that, the condition of preservation the positivity of the density matrix is automatically satisfied, as for any positive operator P , KUK^\dagger is positive as well. Hence the above

transformation is a positive linear map, i.e. it transforms positive operators into positive operators. In fact, it is a *completely positive* (CP) map, which means that even when the map is trivially extended to larger space it remains positive.

The evolution of a quantum state described by Eq. (2.22) is in general no longer unitary, and may in particular generate mixed states out of pure input states, and lead to the so called decoherence of quantum states. To see it, let us go back to the Stern-Gerlach example. Immediately, after the interaction the output state is given by Eq. (2.6). Let us calculate the corresponding reduced density matrix of the S (spin) system:

$$\rho_S(\delta t) = \text{Tr}_M (|\psi(\delta t)\rangle\langle\psi(\delta t)|) = \begin{bmatrix} |c_+|^2 & c_+^* c_- \langle\phi_-|\phi_+\rangle \\ c_-^* c_+ \langle\phi_+|\phi_-\rangle & |c_-|^2 \end{bmatrix}, \quad (2.23)$$

where M now corresponds to the spatial degree of freedom, and the reduced density matrix is written in the $\{|+\rangle, |-\rangle\}$ basis. Comparing the above formula, with the density matrix of the input spin state:

$$|s\rangle\langle s| = \begin{bmatrix} |c_+|^2 & c_+^* c_- \\ c_-^* c_+ & |c_-|^2 \end{bmatrix} \quad (2.24)$$

we see that while the diagonal elements remain unchanged, the off-diagonal elements are being suppressed the more the more orthogonal (distinguishable) are states $|\phi_\pm\rangle$ —the process which we call decoherence. In other words, the more the environment (in this case spatial degrees of freedom) get the information on the spin state of the system the stronger the resulting decoherence process. Note, that decoherence process distinguishes a preferred basis, of the so-called pointer states (in this case $|\pm\rangle$ states), which are not affected by the decoherence process, but superposition of these states are affected, and in the extreme case are transformed into mixtures of pointer states.

Problems

Problem 2.1 Reanalyze the Stern-Gerlach experiment as described in Sec. 2.1, but this time allow the particle to freely evolve, after the interaction with the magnetic field, for a time t , after which time the measurement of position of the particle is performed—assume the standard free evolution of a particle with mass m under the $H_{\text{free}} = p^2/2m$ hamiltonian. Write the probability distribution for detecting a particle at a given point z , and derive the corresponding POVM operators for this measurement. Discuss the limit $t \rightarrow \infty$ and compare it with the direct measurement of momentum that was discussed in the main text.

Problem 2.2 A general state of a two-dimensional quantum system (a qubit) can be written as:

$$|\psi_{\theta,\varphi}\rangle = \cos(\theta/2)|0\rangle + \exp(i\varphi)\sin(\theta/2)|1\rangle,$$

where $\theta \in [0, \pi]$, $\varphi \in [0, 2\pi]$ and $|0\rangle, |1\rangle$ form an orthonormal basis. As a result, we can imagine a state of a qubit as a point on a sphere—the so called Bloch sphere.

Consider the following set of operators, parameterized with θ, φ :

$$\Pi_{\theta,\varphi} = c|\psi_{\theta,\varphi}\rangle\langle\psi_{\theta,\varphi}|,$$

where c is some normalization constant independent of θ, φ .

- Is the set of operators $\Pi_{\theta,\varphi}$ a legitimate POVM? If yes, what is the value of the c constant (assume a standard integration measure on the sphere $d\theta d\varphi \sin\theta$).
- Apply the above generalized measurement, and calculate the corresponding probability distribution as a function of (θ, φ) if the state that was measured was $|0\rangle$. This distribution can be treated as representation of information on how well we can identify a given a state on the Bloch sphere if we have one copy at our disposal.

Problem 2.3 Consider the following unitary operation U representing interaction of the qubit (S) with “measuring device” (M):

$$U|m\rangle_S \otimes |0\rangle_M = \frac{1}{2}|m\rangle_S \otimes |0\rangle_M (\sqrt{2-p} + (-1)^m \sqrt{p}) + \frac{1}{2}|m\rangle_S \otimes |1\rangle_M (\sqrt{2-p} - (-1)^m \sqrt{p}),$$

where $m = 0, 1$, and parameter $0 \leq p \leq 1$ represents the “strength of the interaction” between S and M .

- Using Kraus operators, (operators K_i) write down effective evolution of a general state of a qubit S under this interaction, in situation when no particular measurement result is observed in M —we calculate $\rho'_S = \sum_i K_i \rho_S K_i^\dagger$. Interpret the evolution in the language of Bloch sphere transformation, where the general mixed state of a qubit can be parameterized using a three dimensional vector \vec{n} : $\rho = 1/2(\mathbb{1} + \vec{\sigma} \cdot \vec{n})$, where $|\vec{n}| \leq 1$, and $\vec{\sigma}$ is a vector consisting of Pauli matrices
- Write down measurement operators Π_0, Π_1 acting on system S corresponding to projecting the “measuring device” M on states $|0\rangle_M, |1\rangle_M$.
- Consider a general qubit state $|\psi\rangle$ parameterized using angles θ, φ on the Bloch sphere. Write down probabilities of obtaining measurement results that correspond to measurement operators Π_0, Π_1 , and the respective post-measurement states of the qubit S .

Problem 2.4 Consider a generalization of the Stern-Gerlach experiment model, to the case where we want to measure a general observable \hat{A} of an arbitrary dimensional system, with corresponding Hilbert space \mathcal{H} . We again couple the system to the “measurement” device, states of which we describe using Hilbert space \mathcal{H}_M . We assume that initial state of the system+measuring device is given by $|\psi\rangle \otimes |\phi\rangle \in \mathcal{H} \otimes \mathcal{H}_M$, where $|\psi\rangle$ is an arbitrary state, while $|\phi\rangle$ is a gaussian state of position uncertainty σ as in the Stern-Gerlach experiment example. We assume that the coupling is described by the following Hamiltonian: $H = g\hat{A} \otimes \hat{p}\delta(t)$, that acts instantaneously at time $t = 0$ and where \hat{p} is the momentum operator on space \mathcal{H}_M . Evolve the state using the above Hamiltonian, and discuss the effects of measuring the position of the measuring device at times $t > 0$. Write down the corresponding POVM operators Π_x and describe the conditions of weak vs. strong measurement regime.

Chapter 3

Uncertainty relations

Figure 3.1: Schemes illustrating different variants of uncertainty relations

3.1 Heisenberg-Robertson uncertainty relation

Let us start by recalling the standard formulation of the uncertainty relation that one may encounter in all quantum mechanics textbooks. Given some observable A one defines its uncertainty on a given state $|\psi\rangle$ as $\Delta A = \sqrt{\langle\psi|A^2|\psi\rangle - \langle\psi|A|\psi\rangle^2}$. Then one proves that for two observables A and B the following inequality holds

$$\Delta A \Delta B \geq \frac{|\langle[A, B]\rangle|}{2}, \quad (3.1)$$

which we will refer to as the Heisenberg-Robertson uncertainty relation. In case of position and momentum operators the uncertainty relation reduces to

$$\Delta x \Delta p \geq \frac{\hbar}{2}. \quad (3.2)$$

In order to understand the physical content of the above inequality, we should keep in mind that both ΔA and ΔB are quantities that one determines via measurement of either A or B observables performed on *different* realizations of state $|\psi\rangle$. There is no notion of precision of measuring one observable vs. disturbance of the other observable on the same physical system, which was actually the

main message of the so called Heisenberg microscope thought experiment. Neither there is any notion of joint measurement of the two quantities on a given quantum state. Here, we simply state that the quantum state $|\psi\rangle$ will lead to the following spread of measurement results if observable A or observable B is measured independently of each other. In short, this relation reflects the inherent uncertainty that is present in the quantum state itself with respect to measuring different physical quantities.

3.2 Precision vs. Disturbance uncertainty relation

To understand the difference between the Heisenberg-Robertson formulation of the uncertainty relation and the actual Heisenberg microscope thought experiment, let us recall the Heisenberg microscope setting. Consider a lens with aperture a at a distance l from the illuminated object. For simplicity we assume that the object is imaged using the lens on a very distant screen, see Fig. ???. We can quantify the precision as a minimal size of a light spot in the object plane that correspond to light travelling at a given angle to the screen. From standard optics considerations one can show, that the precision is then of the order: $\delta x \approx \frac{\lambda}{a} l$. Note that we use a different notation, namely δx rather than Δx , to clearly discriminate between the precision of measurement (δx) vs. the inherent uncertainty of the

physical quantity in the state itself (Δx).

In order to assess the disturbance to the momentum of the particle, note that when we collect light arriving at a given angle on the screen, we cannot say exactly at what angle the photon has been emitted from the particle, as it could travel along different paths through the lens. The angle spread is roughly $\varphi \approx a/l$ (we assume the angle is small), and hence the uncertainty in the x -component of momentum transfer to the particle is $dp \approx \frac{h}{\lambda} \frac{a}{l}$ —note again a different notation dp for the disturbance. So finally we can write $\delta x dp \gtrsim \hbar$.

As the reader might have noticed the above reasoning is very qualitative and in fact it is not an easy task to rigorously define the concept of disturbance of the observable (see e.g. a review paper [?]). Moreover, the approach discussed above, breaks the symmetry between the x and p quantities, which is not always desired. Because of this conceptual difficulties we will not study this approach here in detail, and will rather focus on a symmetric and easier to study from a quantitative perspective concept of joint measurement—see Fig. 3 to see the formal difference between these three approaches.

3.3 Joint measurement

Within the concept of joint measurement, our goal is to obtain information on two quantities in the best possible way even if their corresponding observables do not commute. In particular we will want to obtain a limit on the possibility of joint measurement of position and momentum. Clearly, it will not be possible to obtain simultaneous measurement of x and p equivalent to sharp measurements of both quantities. Still, if we allow for a compromise of performing „smeared” position and momentum measurements we may be able to construct the corresponding joint measurement protocol. We start with a general definition.

Definition 3.1. Π_a and Π_b POVMs are

jointly measurable iff there exist a POVM $\Pi_{a,b}$ such that $\Pi_a = \int db \Pi_{a,b}$, $\Pi_b = \int da \Pi_{a,b}$.

Remark. If Π_x , Π_p commute, we simply take $\Pi_{x,p} = \Pi_x \Pi_p$. If not we cannot make this construction (product will not be hermitian/positive).

In particular, we cannot find a joint measurement operator representing ideal (sharp) position and momentum measurements. But, if we consider smeared position and momentum measurements:

$$\tilde{\Pi}_x = \int dx' \mu(x-x') |x'\rangle \langle x'| \quad (3.3)$$

$$\tilde{\Pi}_p = \int dp' \nu(p-p') |p'\rangle \langle p'|, \quad (3.4)$$

where ν , μ are smearing functions (e.g. gaussians), then maybe it is possible to find $\tilde{\Pi}_{x,p}$ such that $\tilde{\Pi}_x = \int dp \tilde{\Pi}_{x,p}$, $\tilde{\Pi}_p = \int dx \tilde{\Pi}_{x,p}$.

Indeed, consider:

$$\tilde{\Pi}_{x,p} = \frac{1}{2\pi\hbar} D(x,p) \Pi_0 D(x,p)^\dagger, \quad (3.5)$$

where $D(q,p) = e^{\frac{ip\hat{x} - ix\hat{p}}{\hbar}}$ is the displacement operator and $\Pi_0 \geq 0$, $\text{Tr} \Pi_0 = 1$. Equivalently:

$$\tilde{\Pi}_{x,p} = \frac{1}{2\pi\hbar} e^{-\frac{ip\hat{x}}{\hbar}} e^{\frac{ix\hat{p}}{\hbar}} \Pi_0 e^{-\frac{ix\hat{p}}{\hbar}} e^{\frac{ip\hat{x}}{\hbar}}, \quad (3.6)$$

where we have used $e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]}$.

We first check whether the above constructed $\tilde{\Pi}_{x,p}$ is a legal POVM. Positivity is clear, as $D(x,p)$ is unitary and Π_0 is positive. We are therefore left to check the completeness condition. Calculating matrix elements in the position eigenbasis

$$\begin{aligned} \frac{1}{2\pi\hbar} \langle x'| \int dx dp e^{-\frac{ipx'}{\hbar}} e^{\frac{ix\hat{p}}{\hbar}} \Pi_0 e^{-\frac{ix\hat{p}}{\hbar}} e^{\frac{ipx''}{\hbar}} |x''\rangle = \\ \delta(x' - x'') \int dx \langle x | \Pi_0 | x \rangle = \delta(x' - x''), \end{aligned} \quad (3.7)$$

we see that indeed $\int dx dp \tilde{\Pi}_{x,p} = \mathbb{1}$.

Let us now calculate marginal probability distributions of x and p resulting from application of this joint measurement to a general

state ρ . Writing $\Pi_0 = \int dx' dx'' (\Pi_0)_{x''}^{x'} |x'\rangle \langle x''|$ in the position eigebasis, we get

$$\begin{aligned} p(x) &= \text{Tr} \left(\rho \int dp \tilde{\Pi}_{x,p} \right) = \\ &= \frac{1}{2\pi\hbar} \text{Tr} \left(\rho \int dp dx' dx'' (\Pi_0)_{x''}^{x'} |x'+x\rangle \langle x''+x| e^{\frac{ip(x''-x')}{\hbar}} \right) \\ &= \text{Tr} \left(\rho \int dx' (\Pi_0)_{x'-x}^{x'} |x'\rangle \langle x'| \right) = \text{Tr}(\rho \tilde{\Pi}_x), \end{aligned} \quad (3.8)$$

from which we see that indeed:

$$\tilde{\Pi}_x = \int dx' \mu(x' - x) |x'\rangle \langle x'|, \quad (3.9)$$

where $\mu(x) = \langle x | \Pi_0 | x \rangle$ is the smearing function. Analogously we can show that the momentum probability distribution corresponds to

$$\tilde{\Pi}_p = \int dp' \nu(p' - p) |p'\rangle \langle p'|, \quad (3.10)$$

where $\mu(p) = \langle p | \Pi_0 | p \rangle$.

Notice that in fact Π_0 can be regarded as a state ($\Pi_0 \geq 0$, $\text{Tr}(\Pi_0) = 1$.) As such it must satisfy standard Heisenberg-Roberstson uncertainty relation: $\Delta_{\Pi_0} x \Delta_{\Pi_0} p \geq \hbar/2$.

What is the uncertainty relation for the actually measured values of x and p ? $\tilde{\Delta}x \tilde{\Delta}p$, where $\tilde{\Delta}$ should remind us that this are distribution obtained using the joined measurement. Using the properties of the convolution we get:

$$\tilde{\Delta}^2 x = \Delta_{\Pi_0}^2 x + \Delta^2 x, \quad (3.11)$$

where $\Delta^2 x$ is the standard variance of sharp position measurement on ρ . We can write analogous formula for for the momentum distribution. Clearly the final measurement distribution is broadened compared to sharp po-

Figure 3.2: Model of joint position and momentum measurement

sition measurements. Finally we can write

$$\begin{aligned} &= (\Delta_{\Pi_0}^2 x + \Delta^2 x) (\Delta_{\Pi_0}^2 p + \Delta^2 p) \geq \\ &= \frac{\hbar^2}{4} (\Delta_{\Pi_0}^2 x + \Delta^2 x) \left(\frac{1}{\Delta_{\Pi_0}^2 x} + \frac{1}{\Delta^2 x} \right) \geq \hbar^2 \end{aligned} \quad (3.12)$$

where we used the standard Heisenberg-Robertson uncertainty relations for both ρ and Π_0 states, and also the fact that $(a + b)(1/a + 1/b) \geq 4$, for arbitrary $a, b \geq 0$. So finally, we get

$$\tilde{\Delta}x \tilde{\Delta}p \geq \hbar \quad (3.13)$$

and we see that the joint-measurement uncertainty relation differs by a factor of 2 from the standard Heisenberg-Robertson uncertainty relation. Intuitively, this is due to the fact that in final distribution the inherent uncertainty of state we measure is combined together with the uncertainty of the state which is the building block of the measurement itself, and on which we effectively project the measure state.

Moreover, we can interpret, the widths of the smearing functions Δ_{Π_0} , $\Delta_{\Pi_0} p$ as precisions of the actual measurements, and therefore write $\delta x = \Delta_{\Pi_0} x$ and hence: $\delta x \delta p \geq \hbar/4$, where now we have a tradeoff between the precision of measuring x and the precision of measuring p expressed via the familiar uncertainty relation.

Example 3.1 In order to illustrate the above considerations consider the following model of joint position-momentum measurement. Consider a particle S travelling in one dimension, with which we associate position and momentum operators (dimensionless) \hat{x}_S , \hat{p}_S , satisfying $[\hat{x}_S, \hat{p}_S] = i$. Initially the particle is in state $|\psi\rangle_S$. Consider a joint position and momentum measurement where particle S interacts with two “measuring devices” M_1 i M_2 through a unitary evolution:

$$|\Psi\rangle_{SM_1M_2} = U |\psi\rangle_S \otimes |0\rangle_{M_1, M_2}, \quad U = e^{-i(\hat{x}_S \hat{p}_{M_1} - \hat{p}_S \hat{x}_{M_2})}, \quad (3.14)$$

where $|0\rangle_{M_1, M_2}$ is the initial state of the measuring devices. After the action of U , position (x_{M_1}) and momentum (p_{M_2}) is measured of respectively systems M_1 and M_2 (these measurements commute!). As a result of measurement we obtain a certain joint probability distribution of measuring position and momentum $p(x, p)$ on state $|\psi\rangle_S$

We start by evolving \hat{x}_1, \hat{p}_2 in the Heisenberg picture:

$$\hat{x}_1^{\text{out}} = \hat{x}_s + \hat{x}_1 - \frac{1}{2}\hat{x}_2 \quad (3.15)$$

$$\hat{p}_2^{\text{out}} = \hat{p}_s + \hat{p}_2 - \frac{1}{2}\hat{p}_1, \quad (3.16)$$

where we have used $e^A B e^{-A} = B + [A, B] + \frac{1}{2}[A, [A, B]] + \dots$. We can define $\hat{x}_M = \hat{x}_1 - \frac{1}{2}\hat{x}_2$, $\hat{p}_M = \hat{p}_2 - \frac{1}{2}\hat{p}_1$, since $[\hat{x}_M, \hat{p}_M] = i\hbar$. We assume that ${}_{12}\langle 0|\hat{x}_M|0\rangle_{12} = 0$, ${}_{12}\langle 0|\hat{p}_M|0\rangle_{12} = 0$ as thanks to this $\langle \hat{x}_1^{\text{out}} \rangle = \langle x_S \rangle$, $\langle \hat{p}_2^{\text{out}} \rangle = \langle p_S \rangle$. As a result:

$$\Delta^2 x_1^{\text{out}} = \Delta^2 x_{|\psi\rangle} + \Delta^2_{|0\rangle_{12}} x_M \quad (3.17)$$

$$\Delta^2 p_2^{\text{out}} = \Delta^2 p_{|\psi\rangle} + \Delta^2_{|0\rangle_{12}} p_M \quad (3.18)$$

and finally, using the Heisenberg-Robertson uncertainty relations for $|\psi\rangle$ and $|0\rangle_{12}$ we get the uncertainty relation for joint measurement of x and p :

$$\tilde{\Delta}_x \tilde{\Delta}_p = \Delta x_1^{\text{out}} \Delta p_2^{\text{out}} \geq \hbar. \quad (3.19)$$

In order to saturate the above inequality we should choose a state $|0\rangle_{12}$ such that it minimizes the standard Heisenberg-Robertson uncertainty relation for the x_M and p_M quantities.

Problems

Problem 3.1 Using state $|0\rangle_{M_1, M_2}$ found in the example above, prove that the set of POVM operators corresponding to the above described model of joint measurements $\Pi_{x,p}$ [so that the joint distribution of measuring x and p is given as $p(x, p) = \text{Tr}(|\psi\rangle\langle\psi|\Pi_{x,p})$] has the following form:

$$\Pi_{x,p} = \frac{1}{2\pi} |(x, p)\rangle\langle(x, p)|, \quad |(x, p)\rangle = \frac{1}{\pi^{1/4}} \int dx' e^{-\frac{(x'-x)^2}{2}} e^{ipx'} |x'\rangle, \quad (3.20)$$

where $|(x, p)\rangle$ is the so called coherent state with mean value of position and momentum equal x and p respectively. Therefore, we have a nice interpretation of the joined position and momentum measurements as projections on coherent states:

$$p(x, p) = \frac{1}{2\pi} |\langle\psi|(x, p)\rangle|^2 \quad (3.21)$$

Remark: in quantum optics, the above probability distribution is called the Husimi representation.

Part II

Classical estimation theory

The main task in classical estimation theory is to find the optimal way to determine a parameter in a statistical model using the observed data. The statistical model is described via a probability distribution $p_\theta(x)$ yielding probability of a given event x that depends on the value of an unknown parameter θ . This task is most commonly pursued using two different conceptual frameworks.

In one framework the parameter θ to be estimated has a fixed but unknown value. Given the observed event x , an estimator $\tilde{\theta}(x)$ ascribes a given value of parameter depending on the observed event. In such a framework the resulting estimator will be based solely on the observed data, typically from a series of repeated experiments. We will refer to this framework as the *frequentist* approach, and we will discuss it in chapter 4.

The alternative, Bayesian approach, treats the parameter θ itself as a random variable, with some prior distribution $p(\theta)$, while the statistical model provides us with conditional probability $p(x|\theta)$ (notice the change of notation from $p_\theta(x)$ to $p(x|\theta)$ which reflects the fact that θ is now a random variable as well). The Bayesian approach is preferred in situations where the experiment is repeated a small number of times and we have some important prior information on the parameter that we want to incorporate in our estimating process. The Bayesian approach is discussed in chapter 5. Both approaches typically yield equivalent statements in the limit of infinite number of repetitions of the experiment.

Chapter 4

Frequentist approach

We adopt the frequentist approach here, and consider a family of probability distributions $p_\theta(x)$ parameterized by an unknown parameter θ . For simplicity of presentation we first focus on single-parameter estimation and will generalize our results to multi-parameter case in section 4.3.

4.1 Optimal unbiased estimator

In order to provide an intuition into the problem of determining the optimal estimator let us start with a simple example.

Example 4.1 Consider N identically and independently distributed (i.i.d.) random variables: $\mathbf{x} = (x_1, \dots, x_N)$, where $x_i = \theta + w_i$ and $w_i \sim \mathcal{N}(0, \sigma^2)$ is a normally distributed random variable with mean 0 and variance σ^2 . As a result $x_i \sim \mathcal{N}(\theta, \sigma^2)$. More explicitly, we can write the joint probability of observing measurement events \mathbf{x} as

$$p_\theta(\mathbf{x}) = p_\theta(x_1) \cdots p_\theta(x_N), \quad (4.1)$$

where

$$p_\theta(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \theta)^2}{2\sigma^2}}. \quad (4.2)$$

Assume we observe a given sequence of events: (x_1, \dots, x_N) . What will be the optimal way to estimate θ ? Natural guess is that we should take the average $\tilde{\theta}(\mathbf{x}) = \sum_i x_i / N$, but can we prove this is the optimal choice?

In what follows, we will quantify the optimality of an estimator $\tilde{\theta}$ using its mean squared distance from the true value of the

parameter:

$$\Delta^2 \tilde{\theta} = \int dx (\tilde{\theta}(x) - \theta)^2 p_\theta(x). \quad (4.3)$$

Since within the frequentist framework the parameter θ is unknown but fixed, we have to specify some constraints on the class of estimators we will be considering. Notice, that otherwise there is no fundamental limit on precision of estimator as we might define the estimator $\tilde{\theta}(x) = \theta_0$ to be a constant function and if we are lucky and $\theta_0 = \theta$ we have an estimator with zero uncertainty. Of course, it is clear that such estimators are useless in practice. We will therefore require from our estimators to satisfy the *unbiasedness* condition which excludes the above mentioned pathological cases.

Definition 4.1 (Unbiased estimator). We say that an estimator $\tilde{\theta}$ is unbiased, if and only if for all θ :

$$\langle \tilde{\theta} \rangle = \int dx \tilde{\theta}(x) p_\theta(x) = \theta, \quad (4.4)$$

which is equivalent to saying that on average the estimator returns the true value for all values of parameter θ .

The goal of estimation theory can now be formulated as the task of determining the unbiased estimator that provides the minimum variance—the *minimum variance unbiased estimator*. Interestingly it might happen that such an estimator does not exist, in the sense

that there is no single estimator that is optimal for the whole range of parameters θ (see Problem ??).

Recalling that the frequentist approach assumes a fixed but unknown parameter, it is typical that we deal with situation where we know roughly the parameter value to be around some value θ_0 and want to estimate it precisely staying within some small interval around it. It is therefore useful to introduce a weaker condition of local unbiasedness, which will actually be sufficient to derive all the bounds that will follow, and moreover there will be no issue of nonexistence of minimal variance locally unbiased estimator.

Definition 4.2 (locally unbiased estimator). We say that an estimator $\tilde{\theta}$ is locally unbiased at $\theta = \theta_0$, if and only if

$$\langle \tilde{\theta} \rangle_{\theta=\theta_0} = \int dx \tilde{\theta}(x) p_{\theta_0}(x) = \theta_0, \quad (4.5)$$

$$\left. \frac{d\langle \tilde{\theta} \rangle}{d\theta} \right|_{\theta=\theta_0} = \int dx \tilde{\theta}(x) \left. \frac{dp_{\theta_0}(x)}{d\theta} \right|_{\theta=\theta_0} = 1, \quad (4.6)$$

which means that we only expect the estimator to track the true parameter up to the first order around a given value of parameter $\theta = \theta_0$.

Example 4.1 (continued) Considering the same gaussian example as before, we see that indeed the proposed estimator $\tilde{\theta}(\mathbf{x}) = \sum_i x_i/N$ is unbiased, whereas its uncertainty reads:

$$\Delta^2 \tilde{\theta} = \left\langle \left(\frac{1}{N} \sum_i x_i - \theta \right)^2 \right\rangle = \frac{\sigma^2}{N}. \quad (4.7)$$

The question remains if this is the minimal possible variance?

4.2 Cramér-Rao bound

We would like now to derive a lower bound on variance of any unbiased (locally) estimator, the so called Cramér-Rao (CR) bound. Thanks to this once we are able to show that

a given estimator saturates the bound we will be sure that it is optimal.

Theorem 4.1 (Cramér-Rao bound). Let $p_\theta(x)$ be a family of probability distributions. Provided $p_\theta(x)$ satisfies some regularity conditions (see the proof), precision of any locally unbiased estimator $\tilde{\theta}$ is lower bounded by:

$$\Delta^2 \tilde{\theta} \geq \frac{1}{F}, \quad F = \int dx \frac{\dot{p}_\theta(x)^2}{p_\theta(x)}, \quad (4.8)$$

where $\dot{p}_\theta(x) = \frac{dp_\theta(x)}{d\theta}$, and F is called the *Fisher Information* (FI). For simplicity of notation we have replaced θ_0 with θ .

Proof. We assume

$$\int dx \tilde{\theta}(x) \dot{p}_\theta(x) = 1, \quad (4.9)$$

$$\int dx \dot{p}_\theta(x) = 0, \quad (4.10)$$

where the first condition is the local unbiasedness condition, while the second is the formal requirement for regularity of $p_\theta(x)$ (if $p_\theta(x)$ is regular we may enter with the integral under the derivative and trivially satisfy this condition)—see Problem ?? to see an example of the model where this regularity assumption is not satisfied and there is no lower bound on uncertainty of the estimator.

Consider the following chain of inequalities

$$\begin{aligned} \Delta^2 \tilde{\theta} \cdot F &= \int dx p_\theta(x) (\tilde{\theta}(x) - \theta)^2 \cdot \int dx \frac{\dot{p}_\theta(x)^2}{p_\theta(x)} = \\ &= \int dx \left[\sqrt{p_\theta(x)} (\tilde{\theta}(x) - \theta) \right]^2 \cdot \int dx \left(\frac{\dot{p}_\theta(x)}{\sqrt{p_\theta(x)}} \right)^2 \stackrel{\text{C-S}}{\geq} \\ &= \left(\int dx (\tilde{\theta}(x) - \theta) \dot{p}_\theta(x) \right)^2 = 1, \quad (4.11) \end{aligned}$$

where we have used the Cauchy-Schwarz (C-S) inequality and utilized the local unbiasedness and regularity conditions in the last step. ■

Remark. One can encounter different but equivalent formulas for the FI:

$$F = \left\langle \left(\frac{d}{d\theta} \log p_\theta(x) \right)^2 \right\rangle = - \left\langle \frac{d^2}{d\theta^2} \log p_\theta(x) \right\rangle. \quad (4.12)$$

Additivity of FI. The FI is additive for product distributions. Let $p_\theta^{(12)}(x_1, x_2) = p_\theta^{(1)}(x_1) p_\theta^{(2)}(x_2)$, then $F^{(12)} = F^{(1)} + F^{(2)}$.

This is the justification for referring to this quantity as information. In particular, given N i.i.d. random variables x_i , $F^{(N)} = NF$, where F is the FI for single random variable, and in such cases the CR bound yields

$$\Delta^2 \tilde{\theta} \geq \frac{1}{NF}, \quad (4.13)$$

showing the expected $1/N$ decrease in estimation variance as the number of repetitions of experiment increases.

Example 4.1 (continued) Let us calculate the FI for the Gaussian example studied in this chapter. Since we deal with N i.i.d. random variables, we can immediately say that $F^{(N)} = NF$, where F is the FI for the Gaussian $p_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\theta)^2/2\sigma^2}$, and equals $F = 1/\sigma^2$. Hence we obtain

$$\Delta^2 \tilde{\theta} \geq \frac{\sigma^2}{N} \quad (4.14)$$

demonstrating that indeed the estimator considered before is optimal. In general an estimator that saturates the CR bound is called *efficient*.

Condition for saturability of the CR bound. Recalling the derivation of the CR bound, we see that the saturation of the CR bound is equivalent to saturation of the Cauchy-Schwarz inequality which is equivalent to:

$$\lambda(\theta) \sqrt{p_\theta(x)} (\tilde{\theta}(x) - \theta) = \frac{\dot{p}_\theta(x)}{\sqrt{p_\theta(x)}} \quad (4.15)$$

or equivalently

$$\frac{d}{d\theta} \log p_\theta(x) = \lambda(\theta) (\tilde{\theta}(x) - \theta), \quad (4.16)$$

where $\lambda(\theta)$ is arbitrary function. One can check the the above condition indeed holds for the exemplary Gaussian model we discussed in this section, provided we set $\tilde{\theta}(x) = \sum_i x_i/N$, $\lambda(\theta) = N/\sigma^2$.

4.3 Multi-parameter case

We now consider a general situation where we want to estimate multiple parameters $\theta =$

$(\theta_1, \theta_2, \dots, \theta_P)$. The object which is a natural generalization of the estimator variance is the estimator covariance matrix \mathbb{C} :

$$\mathbb{C}_{ij} = \int dx p_\theta(x) (\tilde{\theta}_i(x) - \theta_i) (\tilde{\theta}_j(x) - \theta_j). \quad (4.17)$$

Diagonal elements represent the variances of estimators of a particular parameter, while off-diagonal terms represent potential correlations between estimation of different parameters. The multi-parameter generalization of the CR bound is a matrix inequality bounding the \mathbb{C} matrix with the FI matrix.

Theorem 4.2 (Multi-parameter CR bound).

$$\mathbb{C} \geq \mathbb{F}^{-1}, \quad \mathbb{F}_{ij} = \int dx \frac{\partial_i p_\theta(x) \partial_j p_\theta(x)}{p_\theta(x)}, \quad (4.18)$$

where \mathbb{F} is the FI matrix and ∂_i denote differentiation with respect to θ_i parameter. The above matrix inequality should be understood in the sense that $\mathbb{C} - \mathbb{F}^{-1}$ is a positive semi-definite matrix.

Proof. We assume regularity and local unbiasedness conditions, which in the multiparameter case amount to:

$$\int dx \tilde{\theta}_i(x) \partial_j p_\theta(x) = \delta_{ij}, \quad (4.19)$$

$$\int dx \partial_i p_\theta(x) = 0. \quad (4.20)$$

Let us choose some vectors w and v of length P and write

$$\begin{aligned} w^T \mathbb{C} w \cdot v^T \mathbb{F} v &= \int dx \sum_{ij} w_i p_\theta(x) (\tilde{\theta}_i(x) - \theta) (\tilde{\theta}_j(x) - \theta) w_j \\ &\quad \cdot \int dx' \sum_{i'j'} v_{i'} \frac{\partial_{i'} p_\theta(x') \partial_{j'} p_\theta(x')}{p_\theta(x')} v_{j'} = \\ &= \int dx \sum_i w_i \sqrt{p_\theta(x)} (\tilde{\theta}_i(x) - \theta) \sum_j \sqrt{p_\theta(x)} (\tilde{\theta}_j(x) - \theta) w_j \cdot \\ &\quad \int dx' \sum_{i'} v_{i'} \frac{\partial_{i'} p_\theta(x')}{\sqrt{p_\theta(x')}} \sum_{j'} \frac{\partial_{j'} p_\theta(x')}{\sqrt{p_\theta(x')}} v_{j'} \\ &\stackrel{\text{C-S}}{\geq} \left[\int dx \left(\sum_i w_i (\tilde{\theta}_i(x) - \theta) \right) \left(\sum_{i'} v_{i'} \partial_{i'} p_\theta(x) \right) \right]^2 = \\ &\quad (w^T v)^2, \quad (4.21) \end{aligned}$$

where in the last step we have used the local unbiasedness as well as regularity conditions. Choosing

Figure 4.1: Maximum likelihood estimator

$w = Fv$, we get:

$$v^T \mathbb{F} \mathbb{C} \mathbb{F} v \cdot v^T \mathbb{F} v \geq (v^T \mathbb{F} v)^2, \quad (4.22)$$

$$v^T \mathbb{F} \mathbb{C} \mathbb{F} v \geq v^T \mathbb{F} v. \quad (4.23)$$

Since the above inequality is valid for arbitrary v , this implies

$$\mathbb{F} \mathbb{C} \mathbb{F} \geq \mathbb{F} \Rightarrow \mathbb{C} \geq \mathbb{F}^{-1}, \quad (4.24)$$

where the final result we have obtained by acting on both sides with \mathbb{F}^{-1} . ■

Remark. From the derived bound it follows in particular that: $\Delta^2 \tilde{\theta}_i \geq (\mathbb{F}^{-1})_{ii} \geq (\mathbb{F}_{ii})^{-1}$, and the last inequality is in general strict if \mathbb{F} contains nonzero off-diagonal elements.

To see this consider: $1 = e_i^T \sqrt{\mathbb{F}} \sqrt{\mathbb{F}^{-1}} e_i \stackrel{\text{C-S}}{\leq} e_i^T \mathbb{F} e_o e_i^T \mathbb{F}^{-1} e_i$, where e_i is the basis vector with 1 at i -th position and zeros elsewhere. This inequality leads to $(\mathbb{F}^{-1})_{ii} \geq 1/\mathbb{F}_{ii}$.

4.4 Maximum likelihood estimator

Typically, we will encounter situations when there is no unbiased estimator that strictly saturates the CR bound for the parameter we want to estimate. We are therefore looking for some universally applicable recipe to find a good estimator.

Definition 4.3 (Maximum likelihood (ML) estimator). Given a probabilistic model, $p_\theta(x)$, the ML estimator is defined as:

$$\tilde{\theta}_{\text{ML}}(x) = \operatorname{argmax}_\theta [l_x(\theta)], \quad (4.25)$$

where $l_x(\theta) = p_\theta(x)$ is the likelihood function, for which θ is the argument.

In other words, given observed event x we look for such a parameter θ for which probability $p_\theta(x)$ (or equivalently the likelihood $l_x(\theta)$) is maximal—the event is most likely, see Fig. ???. The position of the maximum corresponds to $\tilde{\theta}_{\text{ML}}(x)$.

Remark. In practice, since $l_x(\theta)$ will often be represented as product of many terms (as in e.g. repeated experiment scenarios), it is much more efficient and stable numerically to maximize $\log[l_x(\theta)]$ (the log-likelihood function), as products will turn into sums, and since the log function is monotonic the position of the maximum will remain unchanged.

Example 4.2 Consider our Gaussian example,

$$l_x(\theta) = p_\theta(\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}^N} \prod_{i=1}^N e^{-(x_i - \theta)/\sigma^2}. \quad (4.26)$$

For a given \mathbf{x} we look for the maximum of l . The condition

$$\frac{d}{d\theta} p_\theta(\mathbf{x}) = 0, \text{ or equivalently } \frac{d}{d\theta} \log p_\theta(\mathbf{x}) = 0 \quad (4.27)$$

implies:

$$\tilde{\theta}_{\text{ML}}(\mathbf{x}) = \frac{1}{N} \sum_i x_i. \quad (4.28)$$

We see that the ML estimator is actually the same simple estimator we have proven before to be efficient.

The above apparent coincidence of the ML estimator and the efficient estimator is a general feature. Note that the CR bound saturability condition (4.16) implies that if we take θ such that $\frac{d}{d\theta} \log p_\theta(\mathbf{x}) = 0$, i.e. extremum of the log likelihood function, then the actual efficient estimator $\tilde{\theta}(\mathbf{x}) = \theta = \tilde{\theta}_{\text{ML}}(\mathbf{x})$ (unless $\lambda(\theta) = 0$, which corresponds to a trivial case of FI equal to zero), so indeed the ML estimator is the efficient estimator.

We will now prove the most important theorem of classical estimation theory, namely the asymptotic efficiency of the ML estimator, which means that the ML estimator will asymptotically saturate the CR bound in a model with large number of identical and independent repetitions of the experiment. We will need the following Lemma:

Lemma 4.1. Let $p_1(x)$, $p_2(x)$ be two probability distributions, then

$$D(p_1|p_2) = \int dx p_1(x) \log \frac{p_1(x)}{p_2(x)} \geq 0, \quad (4.29)$$

where $D(p_1|p_2)$ is called the relative entropy.

Proof. The log function is concave, which means that $\log(\sum_i w_i t_i) \geq \sum_i w_i \log t_i$, for $w_i \geq 0$, $\sum_i w_i = 1$. Setting $t_x = \frac{p_2(x)}{p_1(x)}$, $w_x = p_1(x)$ and utilizing the concavity of the log function we get

$$\int dx p_1(x) \log \frac{p_2(x)}{p_1(x)} \leq \log \int dx p_2(x) = 0, \quad (4.30)$$

which ends the proof. ■

We are now ready to prove the main theorem.

Theorem 4.3. Let

$$p_\theta(\mathbf{x}) = p_\theta(x_1) \cdots p_\theta(x_N), \quad (4.31)$$

represent the joint probability distribution for N independent repetitions of an experiment. The ML estimator will be asymptotically unbiased and efficient in the limit $N \rightarrow \infty$, which formally means:

$$\tilde{\theta}_{\text{ML}} \sim \mathcal{N}\left(\theta_0, \frac{1}{FN}\right), \quad (4.32)$$

where θ_0 is the true value of the parameter and F is the FI corresponding to a single experiment $p_\theta(x_i)$ at $\theta = \theta_0$.

Proof. We start by making some technical assumptions concerning the regularity of $p_\theta(x)$. We assume that that $\log p_\theta(x)$ has derivatives up to order 2 and $\langle \partial_\theta \log p_\theta(x) \rangle = 0$. The proof consists of two parts. First we prove asymptotic unbiasedness and then efficiency.

Asymptotic unbiasedness. Let $\tilde{\theta}$ be an estimator. Let us divide the log-likelihood function at $\tilde{\theta}$ by N :

$$\frac{1}{N} l_{\mathbf{x}}(\tilde{\theta}) = \frac{1}{N} \log p_{\tilde{\theta}}(\mathbf{x}) = \frac{1}{N} \sum_i \log p_{\tilde{\theta}}(x_i). \quad (4.33)$$

By the law of large numbers, for almost every sequence \mathbf{x} , we get

$$\frac{1}{N} l_{\mathbf{x}}(\tilde{\theta}) \xrightarrow{N \rightarrow \infty} \int dx p_{\theta_0}(x) \log p_{\tilde{\theta}}(x) \quad (4.34)$$

where θ_0 is the true value. Using Lemma 4.1 we get

$$\int dx p_{\theta_0}(x) \log p_{\tilde{\theta}}(x) \leq \int dx p_{\theta_0}(x) \log p_{\theta_0}(x). \quad (4.35)$$

This shows that the argument $\tilde{\theta}$ for which we obtain the maximum of $l_{\mathbf{x}}(\theta)$, i.e. the ML estimator, in the asymptotic limit $N \rightarrow \infty$ will correspond to the true value.

Asymptotic efficiency. We will start by invoking the mean value theorem, which states that assuming $\theta_0 < \tilde{\theta}$ (the order here is not important) there exist $\theta_0 \leq \bar{\theta} \leq \tilde{\theta}$ such that:

$$\frac{\left. \frac{d \log p_\theta(\mathbf{x})}{d\theta} \right|_{\theta=\tilde{\theta}} - \left. \frac{d \log p_\theta(\mathbf{x})}{d\theta} \right|_{\theta=\theta_0}}{\tilde{\theta} - \theta_0} = \left. \frac{d^2 \log p_\theta(\mathbf{x})}{d\theta^2} \right|_{\theta=\bar{\theta}}. \quad (4.36)$$

If $\tilde{\theta} = \tilde{\theta}_{\text{ML}}$ then $\left. \frac{d \log p_\theta(\mathbf{x})}{d\theta} \right|_{\theta=\tilde{\theta}_{\text{ML}}} = 0$ therefore we get

$$\left. \frac{d \log p_\theta(\mathbf{x})}{d\theta} \right|_{\theta=\theta_0} = \left. \frac{d^2 \log p_\theta(\mathbf{x})}{d\theta^2} \right|_{\theta=\bar{\theta}} (\theta_0 - \tilde{\theta}_{\text{ML}}). \quad (4.37)$$

Let us now consider:

$$\frac{1}{N} \left. \frac{d^2 \log p_\theta(\mathbf{x})}{d\theta^2} \right|_{\theta=\bar{\theta}} = \frac{1}{N} \sum_i \left. \frac{d^2 \log p_\theta(x_i)}{d\theta^2} \right|_{\theta=\bar{\theta}}. \quad (4.38)$$

We know that $\tilde{\theta}_{\text{ML}} \xrightarrow{N \rightarrow \infty} \theta_0$ and hence $\bar{\theta} \xrightarrow{N \rightarrow \infty} \theta_0$. We can therefore write:

$$\frac{1}{N} \left. \frac{d^2 \log p_\theta(\mathbf{x})}{d\theta^2} \right|_{\theta=\bar{\theta}} \xrightarrow{N \rightarrow \infty} \frac{1}{N} \sum_i \left. \frac{d^2 \log p_\theta(x_i)}{d\theta^2} \right|_{\theta=\theta_0} = \left\langle \left. \frac{d^2 \log p_\theta(x_i)}{d\theta^2} \right|_{\theta=\theta_0} \right\rangle = -F. \quad (4.39)$$

Let us define a random variable ξ , which is a sum of N i.i.d variables, as follows:

$$\xi = \frac{1}{\sqrt{N}} \left. \frac{d \log p_\theta(\mathbf{x})}{d\theta} \right|_{\theta=\theta_0} = \frac{1}{\sqrt{N}} \sum_i \left. \frac{d \log p_\theta(x_i)}{d\theta} \right|_{\theta=\theta_0}. \quad (4.40)$$

Note that $\langle \xi \rangle = 0$, while the second moment reads:

$$\langle \xi^2 \rangle = \left\langle \frac{1}{N} \left(\sum_i \left. \frac{d \log p_\theta(x_i)}{d\theta} \right|_{\theta=\theta_0} \right)^2 \right\rangle = \frac{1}{N} \sum_i \left\langle \left(\left. \frac{d \log p_\theta(x_i)}{d\theta} \right|_{\theta=\theta_0} \right)^2 \right\rangle = F. \quad (4.41)$$

By the central limit theorem this implies that $\xi \sim \mathcal{N}(0, F)$. As a result

$$\tilde{\theta}_{\text{ML}} - \theta_0 = \frac{\left. \frac{d \log p_\theta(\mathbf{x})}{d\theta} \right|_{\theta=\theta_0}}{\left. \frac{d^2 \log p_\theta(\mathbf{x})}{d\theta^2} \right|_{\theta=\bar{\theta}}} \sim \mathcal{N}\left(0, \frac{NF}{N^2 F^2}\right) \quad (4.42)$$

so finally:

$$\tilde{\theta}_{\text{ML}} \sim \mathcal{N}(\theta_0, (NF)^{-1}), \quad (4.43)$$

which shows that asymptotically the maximum likelihood estimator is normally distributed and saturates the CR bound. A priori, it is not clear, however, how large N need to be taken to saturate the bound up

to some give accuracy. This depends on the details of the model. ■

Problems

Problem 4.1 Consider two random variables x_1, x_2 distributed according to:

$$x_1 \sim \mathcal{N}(\theta, \sigma^2) \quad (4.44)$$

$$x_2 \sim \begin{cases} \mathcal{N}(\theta, \sigma^2) & \theta \geq 0 \\ \mathcal{N}(\theta, 2\sigma^2) & \theta < 0 \end{cases} \quad (4.45)$$

where $x \sim \mathcal{N}(\theta, \sigma^2)$ means that x is distributed according to a Gaussian (normal) distribution with mean θ and variation σ^2 . Assuming σ is known and the parameter to estimate is θ , try to prove using the Cramera-Rao bound, that in the region $\theta \geq 0$ the optimal estimator is $\tilde{\theta}(x_1, x_2) = \frac{1}{2}(x_1 + x_2)$, while in region $\theta < 0$ the optimal estimator is $\tilde{\theta}(x_1, x_2) = \frac{1}{3}(2x_1 + x_2)$. In this way you will prove that there is no single estimator minimizing the variance for all θ .

Problem 4.2 Consider a generalization of the Cramer-Rao bound, where instead of estimating the parameter θ itself we want to estimate a value of a function $g(\theta)$. Prove that that if $p_\theta(x)$ is a family of probability distribution then for arbitrary locally unbiased estimator $\tilde{g}(x)$, we have

$$\Delta^2 \tilde{g} \geq \frac{g'(\theta)^2}{F} \quad (4.46)$$

where $g'(\theta) = \frac{dg(\theta)}{d\theta}$ and F is the Fisher information for $p_\theta(x)$.

Problem 4.3 We say that $p_\theta(x)$ belongs to the exponential family of probability distributions if and only if

$$p_\theta(x) = e^{a(\theta)+b(x)+c(\theta)d(x)} \quad (4.47)$$

Prove, that in this case, there is always a function $g(\theta)$ for which there exist an efficient estimator—the estimator that saturates the CR bound. A lot of probability distributions belong to this family (see http://en.wikipedia.org/wiki/Exponential_family).

Problem 4.4 Consider a probabilistic model where we register values of N independent random variables x_n , ($n = 0, \dots, N - 1$), where $x_n \sim \mathcal{N}(an + b, \sigma^2)$ (linear dependence + gaussian noise).

- Write down the Fisher matrix corresponding to the two-parameter estimation problem of estimating a and b parameters.
- Using the CR bound derive a lower bound on the minimal achievable estimation variance of the two parameters: $\Delta \tilde{a}$, $\Delta \tilde{b}$. Which one is „easier” to estimate.
- Try to provide estimators saturating the bound—check if by chance these are the same estimators that one uses in the heuristic minimum squared distance method . . .

Problem 4.5 Consider N i.i.d binary valued random variables $x_i \in \{0, 1\}$ ($i = 0, \dots, N - 1$), where $p(x_i = 0) = p$, $p(x_i = 1) = 1 - p$. Consider the problem of estimating parameter p . (Hint: To simplify further calculations, note that what is really relevant in the observed events is the number a zeros and ones in N realizations and not the order in which they appeared).

- What does the Cramér-Rao (CR) bound tells us concerning the best achievable precision of estimating p ?
- Is CR bound saturable for finite N ? What is the optimal estimator?
- Does this family of probability distributions belong to the so called exponential family (see Problem 2 in Problem set 3)?
- Imagine, that in fact $p = \sin^2(\theta/2)$, where $\theta \in [0, \pi]$ and we are actually interested in estimating θ , and not p itself. Derive the CR bound for estimating θ .

e) This time, there is no estimator that saturates the CR bound (check it) for finite N . We can, however, try to use the maximum-likelihood (ML) estimator in order to estimate θ and check whether we can approach the CR bound in the limit of large number of experiment repetitions. Proceed as follows:

- Write a program, generating N i.i.d. realizations of random variable x_i , such that $p(x_i = 0) = \sin^2(\theta/2)$, $p(x_i = 1) = \cos^2(\theta/2)$, for some fixed θ (e.g. $\pi/3$, $\pi/2$, $2/3\pi$) and some fixed N (e.g. $N = 10$). Such a sample of N numbers we will call a single realization of the experiment.
- Generate data for k ($k \approx 1000$, or more) experiments
- For each experiment, find the ML estimator $\tilde{\theta}_{ML}$
- Plot a histogram of obtained values of ML estimator and calculate the spread of the results (standard deviation)— this will be a good approximation of the estimator uncertainty $\Delta\tilde{\theta}$. Compare with the CR bound.
- Repeat above steps for different N , e.g. in the range of 1 to 10000 (of course not for all N but only some representative ones). Generate a plot: estimator uncertainty vs. N and compare it with the CR bound to draw a conclusion concerning the regime where we can claim asymptotic saturation of the CR bound (e.g. you can assume a criterion, that we look for such an N when we are within 1% from the CR bound). Hint: For clarity, it is better to plot $\Delta\tilde{\theta}\sqrt{N}$, rather than $\Delta\tilde{\theta}$, and compare with CR bound for a single realization.

Chapter 5

Bayesian approach

In Bayesian approach, we will write $p(x|\theta)$ instead of $p_\theta(x)$, which is to reflect the fact that θ should be regarded as a random variable itself and not a fixed but unknown parameter as in the frequentist approach. Within the Bayesian approach, apart from $p(x|\theta)$, we need to specify prior distribution $p(\theta)$, which reflects our knowledge of the parameter which we have prior to performing any experiment.

In frequentist approach our goal was to minimize the estimation variance, as given in Eq. (4.3), with local unbiasedness condition imposed. In Bayesian approach the goal is to minimize the average variance:

$$\overline{\Delta^2 \tilde{\theta}} = \int d\theta p(\theta) \int dx (\tilde{\theta}(x) - \theta)^2 p(x|\theta). \quad (5.1)$$

In this case there is no need to impose any additional requirements such as unbiasedness. We simply look for an estimator $\tilde{\theta}$ that minimizes the above quantity.

5.1 Optimal Bayesian estimator

Let us rewrite the formula for the average variance, using the Bayes rule $p(x|\theta)p(\theta) = p(\theta|x)p(x)$, as follows:

$$\overline{\Delta^2 \tilde{\theta}} = \int dx p(x) \int d\theta, (\tilde{\theta}(x) - \theta)^2 p(\theta|x). \quad (5.2)$$

Since $p(x) \geq 0$, and $\tilde{\theta}(x)$ for different x can be treated as independent variables, min-

imization over $\tilde{\theta}$ amounts to minimization of $\int d\theta, (\tilde{\theta}(x) - \theta)^2 p(\theta|x)$ quantity independently for each x over $\tilde{\theta}(x)$. This is a quadratic function in $\tilde{\theta}(x)$ and hence minimization is straightforward, as

$$\frac{d}{d\tilde{\theta}(x)} \int d\theta (\tilde{\theta}(x) - \theta)^2 p(\theta|x) = 0 \quad (5.3)$$

implies

$$\tilde{\theta}(x) = \int d\theta p(\theta|x) \theta = \langle \theta \rangle_{p(\theta|x)}. \quad (5.4)$$

Hence the optimal Bayesian estimator corresponds to the the mean of the posteriori distribution $p(\theta|x)$. The corresponding minimal cost reads:

$$\begin{aligned} \overline{\Delta^2 \tilde{\theta}} &= \int dx p(x) \int d\theta, (\langle \theta \rangle_{p(\theta|x)} - \theta)^2 p(\theta|x) = \\ &= \int dx p(x) \Delta^2 \theta|_{p(\theta|x)} \end{aligned} \quad (5.5)$$

and amount to the average variance of the posteriori distribution.

It is therefore clear that the fundamental object in the Bayesian approach is the posteriori distribution $p(\theta|x)$. This can be calculated

via Bayes rule:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}. \quad (5.6)$$

Note, however, that $p(x)$ is not explicitly given, and calculating it requires performing the following integral: $p(x) = \int d\theta p(x|\theta)p(\theta)$.

$p(x)$ plays a role of a normalization factor for the distribution, while the θ dependence is determined by the product of $p(x|\theta)$ and $p(\theta)$. To get a better intuitive understanding of the Bayesian approach, observe that $p(\theta)$ represents just the prior knowledge while all the information that we get from the data is captured by $p(x|\theta)$. Analyzing these two function one may easily understand what is the relative role of the prior information vs data. If $p(\theta)$ varies much slower with θ compared to $p(x|\theta)$ it means that the prior is largely irrelevant. In the opposite case the the prior dominates our inference strategy.

Example 5.1 Let us again reconsider the gaussian estimation model, where our observations are modeled as N i.i.d. random variables $x_i \sim \mathcal{N}(\theta, \sigma^2)$, from which we want to estimate θ . However, this time we further assume that we have a gaussian prior distribution of the θ parameter itself $\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$. For the model presented we want to find the optimal Bayesian estimator and the resulting estimation uncertainty.

For this model:

$$p(\mathbf{x}|\theta)p(\theta) = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^N e^{-\frac{1}{2}G_{\mathbf{x}}(\theta)}, \quad (5.7)$$

where $G_{\mathbf{x}}(\theta) = \frac{1}{\sigma_\theta^2}(\theta - \mu_\theta)^2 + \frac{1}{\sigma^2} \sum_i (x_i - \theta)^2$.

We immediately see that the posteriori distribution $p(\theta|\mathbf{x}) \sim p(\mathbf{x}|\theta)p(\theta)$ will also be Gaussian. As a result we can easily normalize it and arrive at the final form of the posteriori distribution:

$$p(\theta|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma_{\theta|\mathbf{x}}^2}} e^{-\frac{1}{2\sigma_{\theta|\mathbf{x}}^2}(\theta - \mu_{\theta|\mathbf{x}})^2}, \quad (5.8)$$

where

$$\sigma_{\theta|\mathbf{x}}^2 = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_\theta^2} \right)^{-1}, \quad (5.9)$$

$$\mu_{\theta|\mathbf{x}} = \left(\frac{\sum_i x_i}{\sigma^2} + \frac{\mu_\theta}{\sigma_\theta^2} \right) \sigma_{\theta|\mathbf{x}}^2 \quad (5.10)$$

are respectively the variance and the mean of the posteriori distribution.

The optimal Bayesian estimator which is the mean

of the posteriori distribution $\mu_{\theta|\mathbf{x}}$ may be rewritten in a more appealing form

$$\tilde{\theta}(\mathbf{x}) = \alpha \bar{x} + (1 - \alpha)\mu_\theta, \quad (5.11)$$

where $\bar{x} = \sum_i x_i/N$, $\alpha = \frac{N}{\sigma^2} / \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_\theta^2} \right)$. The above form clearly shows that the optimal Bayesian estimator arises as a result of compromise between what the data suggest (in this case the mean of observed values) and the prior information (in this case the mean of the prior μ_θ) and α represents the weight of the information part.

According to (5.5) the resulting cost will be the average of the posteriori variance. In our model the variance of the posteriori distribution does not depend on \mathbf{x} and hence we may immediately write:

$$\Delta^2 \tilde{\theta} = \sigma_{\theta|\mathbf{x}}^2 = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_\theta^2} \right)^{-1}. \quad (5.12)$$

In the limit $N \rightarrow \infty$ the role of the prior becomes irrelevant and as a result $\tilde{\theta}(\mathbf{x}) \rightarrow \bar{x}$ and $\Delta^2 \tilde{\theta} \rightarrow \sigma^2/N$, which coincides with the results obtained within the frequentist approach.

5.2 Bayesian Cramér-Rao bound

Even though, unlike in the frequentist approach, the recipe for the optimal Bayesian estimator is explicit, it may still be useful to have an easily computable lower bound on achievable estimation uncertainty within the Bayesian framework in the spirit of the Cramér-Rao bound derived within the frequentist approach.

Theorem 5.1 (Bayesian Cramér-Rao bound—van Trees inequality). Given a Bayesian estimation problem with $p(\theta)$ as priori distribution and $p(x|\theta)$ as conditional distribution for observation of an event x , than assuming standard regularity conditions and vanishing of the prior on the ends of the interval over which parameter θ is considered ($\theta \in [\theta_-, \theta_+]$), the following bound on the average variance holds irrespectively of the estimator function chosen

$$\overline{\Delta^2 \tilde{\theta}} \geq \frac{1}{\bar{F} + I}, \quad (5.13)$$

where $\bar{F} = \int d\theta F(\theta)$ is the FI averaged with the prior distribution, while $I =$

$\int d\theta \frac{1}{p(\theta)} \left(\frac{dp(\theta)}{d\theta} \right)^2$ represents the information contribution coming from the prior distribution.

Proof. Let us define two functions

$$f(\theta, x) = \sqrt{p(x|\theta)p(\theta)}(\tilde{\theta}(x) - \theta), \quad (5.14)$$

$$g(\theta, x) = \frac{1}{\sqrt{p(x|\theta)p(\theta)}} \frac{dp(x|\theta)p(\theta)}{d\theta}. \quad (5.15)$$

First observe that $\Delta^2 \tilde{\theta} = \int d\theta dx f(\theta, x)^2$, so the average variance may be viewed as the squared norm of the the function f . Further note:

$$\begin{aligned} \int d\theta dx g(\theta, x)^2 &= \int d\theta dx \frac{p(\theta)}{p(x|\theta)} \left(\frac{dp(x|\theta)}{d\theta} \right)^2 + \\ &\quad \frac{p(x|\theta)}{p(\theta)} \left(\frac{dp(\theta)}{d\theta} \right)^2 + 2 \frac{dp(x|\theta)}{d\theta} \frac{dp(\theta)}{d\theta} = \\ &= \int d\theta p(\theta) F(\theta) + \int d\theta \frac{1}{p(\theta)} \left(\frac{dp(\theta)}{d\theta} \right)^2 = \\ &\quad \bar{F} + I, \end{aligned} \quad (5.16)$$

where we have used regularity assumptions thanks to which $\int d\theta dx \frac{dp(x|\theta)}{d\theta} \frac{dp(\theta)}{d\theta} = 0$. Moreover:

$$\begin{aligned} \int d\theta dx f(\theta, x)g(\theta, x) &= \\ &= \int d\theta dx (\tilde{\theta}(x) - \theta) \frac{dp(x|\theta)p(\theta)}{d\theta} = \\ &= \int dx p(x|\theta)p(\theta)|_{\theta_-}^{\theta_+} - \int d\theta \theta \frac{dp(\theta)}{d\theta} = \\ &= -\theta p(\theta)|_{\theta_-}^{\theta_+} + \int d\theta p(\theta) = 1, \end{aligned} \quad (5.17)$$

where in the last step we performed integration by parts, and we have used the fact that $p(\theta_+) = p(\theta_-) = 0$. Applying now the Cauchy-Schwarz inequality

$$\begin{aligned} \int d\theta dx f(\theta, x)^2 \int d\theta' dx' g(\theta', x')^2 &\geq \\ &\left(\int dx d\theta f(\theta, x)g(\theta, x) \right)^2, \end{aligned} \quad (5.18)$$

we prove the theorem. ■

The above inequality clearly illuminates the role of the data and the prior in Bayesian inference. The \bar{F} quantity corresponds to the information coming from data while I represents the information due to prior. In the large number of experiment repetition limit, we expect \bar{F} to grow linearly with number of repetitions, while I remains constant. Hence in this limit we will recover standard CR inequality

as I will be negligible compared to \bar{F} .

Example 5.1 (continued) For the gaussian Bayesian model considered before, we may calculate the quantities appearing in the Bayesian Cramér-Rao inequality. FI does not depend on θ and hence $\bar{F} = N/\sigma^2$, while $I = 1/\sigma_\theta^2$. We see that the Bayesian Cramer-Rao inequality implies:

$$\overline{\Delta^2 \tilde{\theta}} \geq \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_\theta^2} \right)^{-1}, \quad (5.19)$$

which is exactly the precision achieved by the optimal Bayesian estimator. In this case the inequality is saturated.

[BAYESIAN MULTI-PARAMETER CR BOUND] [JEFFREYS PRIOR]

Problems

Problem 5.1 Consider a Bayesian estimation problem, but with a different cost function than the mean squared error. In case when we want to estimate a phase (or some other angle-like parameter) $\theta \in [0, 2\pi]$, a more practical cost function is a function of the form $C(\theta, \tilde{\theta}) = 4 \sin^2 \left(\frac{\theta - \tilde{\theta}}{2} \right)$, which for small deviations between θ and $\tilde{\theta}$ is equivalent to the variance but respects that fact, that the 2π difference is not relevant. Average cost is then given by:

$$\bar{C} = \int d\theta dx 4 \sin^2 \left(\frac{\theta - \tilde{\theta}(x)}{2} \right) p(x|\theta)p(\theta). \quad (5.20)$$

Find the optimal Bayesian estimator for this cost function.

Problem 5.2 Analyze the conditions for saturation of the Bayesian Cramér-Rao inequality and check if the gaussian model considered during the lecture is the only one for which the inequality is actually saturated.

Part III

Quantum estimation theory

In classical estimation theory, our fundamental object was the probabilistic model given in terms of family of probability distributions $p_\theta(x)$ (or conditional probability $p(x|\theta)$ and prior $p(\theta)$ in the Bayesian approach), and the goal was to find the optimal inference strategy for θ based on observations x . In quantum estimation theory our basic object will be a family of quantum states ρ_θ where θ is the parameter we want to estimate. In order to obtain the family of probability distributions we will additionally need to choose a measurement. If we choose a measurement $\{\Pi_x\}$, then the resulting family of probability distributions are given as $p_\theta(x) = \text{Tr}(\rho_\theta \Pi_x)$ and from now on we can apply the known tools of classical estimation theory.

We see that effectively quantum estimation theory amounts to classical estimation theory + choice of measurement. It will be important to know how to choose the optimal measurement since the amount of accessible information on θ will strongly depend on this choice.

Chapter 6

Quantum frequentist estimation

In order to better grasp the intuitions behind the theory that will be developed in this chapter let us be guided by the following simple example.

Example 6.1 Consider a single qubit system, and the family of states

$$|\psi_\theta\rangle = \frac{1}{\sqrt{2}} (|0\rangle + e^{i\theta}|1\rangle), \quad (6.1)$$

parameterized by an angle $\theta \in [0, 2\pi]$. In the Bloch sphere picture these states correspond to states on the equator. Assume we are given N copies of the state so that

$$\rho_\theta = (|\psi_\theta\rangle\langle\psi_\theta|)^{\otimes N}. \quad (6.2)$$

We want to know how to choose the optimal measurement and estimator in order to estimate θ with the lowest uncertainty possible. Consider two exemplary measurements, corresponding to the following choices of basis, a) $\{|0\rangle, |1\rangle\}$, b) $\{|+\rangle, |-\rangle\}$, where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. In case of measurement a) we see that $p_\theta(0) = 1/2$, $p_\theta(1) = 1/2$, so that the measurements results do not carry any information on the parameter θ . On the other hand, measurement b) leads to probability distributions $p_\theta(\pm) = |\langle\pm|\psi_\theta\rangle|^2 = \frac{1}{2}(1 \pm \cos\theta)$, which seems much more reasonable. We can calculate the correspondin FI for this measurement which yields:

$$F = \frac{1}{p_\theta(+)} \left(\frac{dp_\theta(+)}{d\theta} \right)^2 + \frac{1}{p_\theta(-)} \left(\frac{dp_\theta(-)}{d\theta} \right)^2 = 1. \quad (6.3)$$

As a result given N copies, $F^{(N)} = N$, and the CR bound implies that $\Delta^2\tilde{\theta} \geq 1/N$.

The question is whether this measurement is optimal, or maybe some other measurement could result in higher FI? Note that we have just considered measurements on a single qubit, while in principle given N copies one could consider also collective measurements on all N copies simultaneously. Might that be helpful? In the following sections we will develop tools that will allow us to answer these questions.

In this chapter we will pursue the frequentist approach and try to find a fundamental lower bound on achievable estimation uncertainty. The following theorem is a generalization of the classical Cramér-Rao bound and we first focus on the single-parameter case.

6.1 Quantum Cramér-Rao bound

Theorem 6.1 (Quantum Cramér-Rao bound). Given a family of states ρ_θ , arbitrary measurements and locally unbiased estimators the estimation variance is lower bounded by:

$$\Delta^2\tilde{\theta} \geq \frac{1}{F_Q}, \quad (6.4)$$

where F_Q is the quantum Fisher information (QFI) that is defined as

$$F_Q = \text{Tr}(\rho_\theta \Lambda_\theta^2), \quad (6.5)$$

where Λ_θ is the symmetric logarithmic derivative (SLD) operator defined implicitly via the following equation

$$\frac{d\rho_\theta}{d\theta} = \frac{1}{2} (\Lambda_\theta \rho_\theta + \rho_\theta \Lambda_\theta). \quad (6.6)$$

Proof. First note, that the SLD operator defined in the theorem can be written explicitly if one consider it in the ρ_θ eigenbasis. Let $\rho_\theta = \sum_i \lambda_i |e_i\rangle\langle e_i|$ then

according to defining equation for SLD we have:

$$\begin{aligned} \langle e_i | \frac{d\rho_\theta}{d\theta} | e_j \rangle &= \frac{1}{2} \langle e_i | (\rho_\theta \Lambda_\theta + \Lambda_\theta \rho_\theta) | e_j \rangle = \\ &= \frac{1}{2} (\lambda_i \langle e_i | \Lambda_\theta | e_j \rangle + \langle e_i | \Lambda_\theta | e_j \rangle \lambda_j). \end{aligned} \quad (6.7)$$

As a result we get an explicit formula for the SLD operator:

$$\langle e_i | \Lambda_\theta | e_j \rangle = \frac{2 \langle e_i | \frac{d\rho_\theta}{d\theta} | e_j \rangle}{\lambda_i + \lambda_j}. \quad (6.8)$$

Since ρ_θ is hermitian, then from the above formula it in particular follows that Λ_θ is also a hermitian operator.

Consider a measurement $\{\Pi_x\}$ and the corresponding probability distribution $p_\theta(x) = \text{Tr}(\rho_\theta \Pi_x)$. We want to derive an upper bound on the resulting Fisher information valid for arbitrary measurements:

$$\begin{aligned} F &= \int dx \frac{\left[\text{Tr}(\Pi_x \frac{d\rho_\theta}{d\theta}) \right]^2}{\text{Tr}(\Pi_x \rho_\theta)} = \\ &= \int dx \frac{\left[\text{Tr} \left[\frac{1}{2} \Pi_x (\Lambda_\theta \rho_\theta + \rho_\theta \Lambda_\theta) \right] \right]^2}{\text{Tr}(\Pi_x \rho_\theta)}. \end{aligned} \quad (6.9)$$

Let us just focus on the term in the numerator. Since all ρ_θ , Π_x , Λ_θ are hermitian we may equivalently write

$$\begin{aligned} \left| \text{Tr} \left[\frac{1}{2} \Pi_x (\Lambda_\theta \rho_\theta + \rho_\theta \Lambda_\theta) \right] \right| &= \\ &= \left| \text{Tr} \left(\frac{1}{2} [\Pi_x \Lambda_\theta \rho_\theta + (\Pi_x \Lambda_\theta \rho_\theta)^\dagger] \right) \right| = \\ &= |\text{Re} \text{Tr}(\Pi_x \Lambda_\theta \rho_\theta)| \leq |\text{Tr}(\Pi_x \Lambda_\theta \rho_\theta)|. \end{aligned} \quad (6.10)$$

We now make use of the Cauchy-Schwarz inequality with respect to the Hilbert-Schmidt matrix scalar product:

$$\left| \text{Tr}(AB^\dagger) \right|^2 \leq \text{Tr}(A^\dagger A) \text{Tr}(B^\dagger B), \quad (6.11)$$

where we set $A = \sqrt{\Pi_x} \sqrt{\rho_\theta}$, $B = \sqrt{\Pi_x} \Lambda_\theta \sqrt{\rho_\theta}$ and obtain:

$$|\text{Tr}(\Pi_x \Lambda_\theta \rho_\theta)|^2 \leq \text{Tr}(\rho_\theta \Pi_x) \text{Tr}(\sqrt{\rho_\theta} \Lambda_\theta \Pi_x \Lambda_\theta \sqrt{\rho_\theta}). \quad (6.12)$$

Substituting this inequality to (6.10) and (6.9) we finally arrive at:

$$F \leq \int dx \text{Tr}(\rho_\theta \Lambda_\theta \Pi_x \Lambda_\theta) = \text{Tr}(\rho_\theta \Lambda_\theta^2) = F_Q, \quad (6.13)$$

where we have made use of the completeness property of measurement operators Π_x . This way we have proved that whatever measurement is chosen $F \leq F_Q$. Making use of the classical CR bound we therefore obtain

$$\Delta^2 \tilde{\theta} \geq \frac{1}{F} \geq \frac{1}{F_Q}. \quad (6.14)$$

■

Remark. Note that in the classical case,

$$\frac{dp_\theta(x)}{d\theta} = \frac{d \log p_\theta(x)}{d\theta} p_\theta(x) \quad (6.15)$$

so the object that multiplies the probability distribution and yields its derivative is the logarithmic derivative. The Λ_θ is therefore the operator analog of the logarithmic derivative. Due to non-commutativity this choice is not unique and hence the name SLD indicates that we define it in a symmetric way. It is possible to define e.g. right logarithmic derivative (RLD) via $\frac{d\rho_\theta}{d\theta} = \rho_\theta(x) \Lambda_\theta^R$, and the derivation of the CR bound will also be valid. Still in general the RLD does not exist (notice that it will only exist if the kernel of $\frac{d\rho_\theta}{d\theta}$ is the same as the kernel of $\rho_\theta(x)$). Moreover, even if it exists, it may be shown that in the single parameter case the resulting bound is never tighter than the one based on the SLD. In multi-parameter case, however, it might happen that RLD provides a tighter bound, see Sec. 6.4 for more information. **I do not use RLD**

Additivity of the QFI Similarly to the FI the QFI is additive. Consider a family of states of a bipartite system that are products of states of individual systems:

$$\rho_\theta^{(12)} = \rho_\theta^{(1)} \otimes \rho_\theta^{(2)}. \quad (6.16)$$

Let $\Lambda_\theta^{(1)}$, $\Lambda_\theta^{(2)}$ be SLD operators corresponding to $\rho_\theta^{(1)}$ and $\rho_\theta^{(2)}$ respectively. Then:

$$\begin{aligned} \frac{d\rho_\theta^{(12)}}{d\theta} &= \frac{d\rho_\theta^{(1)}}{d\theta} \otimes \rho_\theta^{(2)} + \rho_\theta^{(1)} \otimes \frac{d\rho_\theta^{(2)}}{d\theta} = \\ &= \frac{1}{2} \left(\Lambda_\theta^{(1)} \rho_\theta^{(1)} + \rho_\theta^{(1)} \Lambda_\theta^{(1)} \right) \otimes \rho_\theta^{(2)} + \\ &= \rho_\theta^{(1)} \otimes \frac{1}{2} \left(\Lambda_\theta^{(2)} \rho_\theta^{(2)} + \rho_\theta^{(2)} \Lambda_\theta^{(2)} \right) = \\ &= \frac{1}{2} \left(\Lambda_\theta^{(1)} \otimes \mathbb{1} + \mathbb{1} \otimes \Lambda_\theta^{(2)} \right) \rho_\theta^{(1)} \otimes \rho_\theta^{(2)} + \leftrightarrow, \end{aligned} \quad (6.17)$$

from which we see that the SLD for the joined state equals

$$\Lambda_\theta^{(12)} = \Lambda_\theta^{(1)} \otimes \mathbb{1} + \mathbb{1} \otimes \Lambda_\theta^{(2)}. \quad (6.18)$$

Consequently the QFI reads:

$$\begin{aligned} F_Q^{(12)} &= \text{Tr}(\rho_\theta^{(12)} \Lambda_\theta^{(12)2}) = \\ &= \text{Tr}(\rho_\theta^{(1)} \Lambda_\theta^{(1)2}) + \text{Tr}(\rho_\theta^{(2)} \Lambda_\theta^{(2)2}) + \\ &+ 2\text{Tr}(\rho_\theta^{(1)} \Lambda_\theta^{(1)}) \cdot \text{Tr}(\rho_\theta^{(2)} \Lambda_\theta^{(2)}) = \\ &F_Q^{(1)} + F_Q^{(2)}, \end{aligned} \quad (6.19)$$

where we have used the fact that $\text{Tr}(\rho_\theta \Lambda_\theta) = \text{Tr} \frac{d\rho_\theta}{d\theta} = 0$.

In particular, when we consider N copies of a quantum state $\rho_\theta^{\otimes N}$, the resulting QFI reads $F_Q^{(N)} = NF_Q$, where F_Q is the QFI corresponding to the single state ρ_θ .

Pure state case Consider a special case where the states in which the parameter is encoded are pure, $\rho_\theta = |\psi_\theta\rangle\langle\psi_\theta|$. In this case the SLD operator may be written explicitly as

$$\Lambda_\theta = 2 \left(|\dot{\psi}_\theta\rangle\langle\psi_\theta| + |\psi_\theta\rangle\langle\dot{\psi}_\theta| \right), \quad (6.20)$$

where $|\dot{\psi}_\theta\rangle = \frac{d|\psi_\theta\rangle}{d\theta}$. Let us check this:

$$\begin{aligned} \frac{1}{2} (\Lambda_\theta |\psi_\theta\rangle\langle\psi_\theta| + |\psi_\theta\rangle\langle\psi_\theta| \Lambda_\theta) &= \\ |\dot{\psi}_\theta\rangle\langle\psi_\theta| + |\psi_\theta\rangle\langle\dot{\psi}_\theta| + (\langle\dot{\psi}_\theta|\psi_\theta\rangle + \langle\psi_\theta|\dot{\psi}_\theta\rangle) |\psi_\theta\rangle\langle\psi_\theta| &= \\ = \frac{d|\psi_\theta\rangle\langle\psi_\theta|}{d\theta}, \end{aligned} \quad (6.21)$$

where we have used the identity $0 = \frac{d\langle\psi_\theta|\psi_\theta\rangle}{d\theta} = \langle\dot{\psi}_\theta|\psi_\theta\rangle + \langle\psi_\theta|\dot{\psi}_\theta\rangle$.

The resulting QFI reads:

$$\begin{aligned} F_Q &= \text{Tr}(|\psi_\theta\rangle\langle\psi_\theta| \Lambda_\theta^2) = \langle\psi_\theta| \Lambda_\theta^2 |\psi_\theta\rangle = \\ &= 4 \left(\langle\psi_\theta|\dot{\psi}_\theta\rangle^2 + \langle\dot{\psi}_\theta|\psi_\theta\rangle^2 + \langle\dot{\psi}_\theta|\dot{\psi}_\theta\rangle + |\langle\psi_\theta|\dot{\psi}_\theta\rangle|^2 \right). \end{aligned} \quad (6.22)$$

Since $\langle\dot{\psi}_\theta|\psi_\theta\rangle + \langle\psi_\theta|\dot{\psi}_\theta\rangle = 0$, if we square it,

we obtain the following identity:

$$\langle\psi_\theta|\dot{\psi}_\theta\rangle^2 + \langle\dot{\psi}_\theta|\psi_\theta\rangle^2 + |\langle\psi_\theta|\dot{\psi}_\theta\rangle|^2 = -|\langle\dot{\psi}_\theta|\psi_\theta\rangle|^2. \quad (6.23)$$

Substituting this to (6.22) we finally arrive at:

$$F_Q = 4 \left(\langle\dot{\psi}_\theta|\dot{\psi}_\theta\rangle - |\langle\dot{\psi}_\theta|\psi_\theta\rangle|^2 \right). \quad (6.24)$$

The above formula has a very intuitive interpretation. There is more information on the parameter accessible in the state the bigger is the derivative $|\dot{\psi}_\theta\rangle$. Still, since states are anyway normalized, the real change has to happen in the direction perpendicular to the state itself, and that is why we need to substitute the component representing the change in the direction of the state itself.

Note, that while in the derivation we have used an explicit formula for the SLD, the SLD for pure states is not a unique operator. In general only for full rank state the SLD is uniquely defined via (6.7), otherwise we may always add terms that are outside of the support of the ρ_θ and this will still lead to the correct formula for the derivative.

Remark (Time-Energy uncertainty relation). Consider the problem of estimating the time of evolution of a quantum state evolving under a known Hamiltonian. Formally consider the family of states:

$$|\psi_t\rangle = e^{-iHt/\hbar} |\psi\rangle, \quad (6.25)$$

where now it is time t which is the parameter to be estimated. As this is a pure state model we can easily calculate QFI using [?] and get

$$F_Q = \frac{4}{\hbar^2} \left(\langle\psi_t| H^2 |\psi_t\rangle - |\langle\psi(t)| H |\psi(t)\rangle|^2 \right), \quad (6.26)$$

which is proportional to the variance of the Hamiltonian on the state. The QCR bound now takes the form

$$\Delta^2 \tilde{t} \Delta^2 H \geq \frac{\hbar^2}{4}, \quad (6.27)$$

which may be viewed as a formal statement

of the time-energy uncertainty relation. The fact that there is no time operator in quantum mechanics does not cause any problem here, since $\Delta^2 \tilde{t}$ is the variance of an estimator and not of an operator.

Example 6.1 (continued) Let us calculate the QFI information for the qubit model. First note that $|\dot{\psi}_\theta\rangle = ie^{i\theta}|1\rangle/\sqrt{2}$, hence

$$F_Q = 4 \left(\frac{1}{2} - \frac{1}{4} \right) = 1. \quad (6.28)$$

Given N copies the QFI equals $F_Q^{(N)} = N$, and hence the QCR bound implies that $\Delta^2 \tilde{\theta} = 1/N$. Note that this is the same value we have obtained, when we calculated FI for the measurement in $|\pm\rangle$ basis. It implies that this measurement is indeed optimal (one can check that actually any measurement in the basis where vectors lie in the equatorial plane of the Bloch sphere will be optimal, so this choice was not unique).

This observation has some far reaching consequences. As a by product we have also proven, that collective measurements are not necessary to achieve the optimal precision—note that QFI for N copies $\rho_\theta^{\otimes N}$ is just N times QFI for a single copy, and hence we can find a measurement on a single copy that makes the corresponding FI equal to QFI of the state, it implies that if we repeat the measurement on N copies we will get N times larger FI, and as result the same value as QFI for the N copy state—note that the derivation of the quantum CR bound allowed for arbitrary measurements, so when considered for the $\rho_\theta^{\otimes N}$ we have taken into account the possibility of collective measurements.

Saturability of the quantum CR bound

In the single qubit example from previous section we have seen that there was a simple measurement for which the corresponding FI was equal to the QFI. Inspecting the derivation of the QCR bound, we see that in order to saturate the Cauchy-Schwarz inequality we need to satisfy

$$\sqrt{\Pi_x} \sqrt{\rho_\theta} = \lambda_x \sqrt{\Pi_x} \Lambda_\theta \sqrt{\rho_\theta}, \quad (6.29)$$

where λ_x is some proportionality constant. Moreover, iff λ is real then inequality (6.10) will also be saturated. This can be seen as

follows:

$$\begin{aligned} |\operatorname{Re} \operatorname{Tr}(\Pi_x \Lambda_\theta \rho_\theta)| &= |\operatorname{Re} \operatorname{Tr}(\sqrt{\rho_\theta} \sqrt{\Pi_x} \sqrt{\Pi_x} \Lambda_\theta \sqrt{\rho_\theta})| = \\ &= |\operatorname{Re} \lambda_x \operatorname{Tr}(\sqrt{\rho_\theta} \Lambda_\theta \Pi_x \Lambda_\theta \sqrt{\rho_\theta})|. \end{aligned} \quad (6.30)$$

Note that the operator under the trace is hermitian so the trace is real. Hence if and only if $\lambda \in \mathbb{R}$ we can remove Re without changing the value of the expression.

Let $\Lambda_\theta = \sum_x l_x |x\rangle\langle x|$ be the eigendecomposition of Λ_θ so that $|x\rangle$ form orthonormal eigenbasis. Now, let us consider a measurement which corresponds to a projection measurement in the eigenbasis of the SLD operator: $\Pi_x = |x\rangle\langle x|$. Note that since this is a projective measurement $\sqrt{\Pi_x} = \Pi_x$. We have:

$$|x\rangle\langle x| \Lambda_x \sqrt{\rho_\theta} = l_x |x\rangle\langle x| \sqrt{\rho_\theta}, \quad (6.31)$$

and hence indeed we satisfy all saturability conditions provided we set $\lambda_x = 1/l_x$.

Remark. Even though we have proven that there always exist a projective measurement for which FI equals to QFI, we need to keep in mind that we still need to satisfy the classical requirement of existence of the estimator that satisfies the classical CR bound in order to claim that actually the QCR bound is saturated. In particular even if there is a single copy measurement for which FI equals the QFI it does not mean we can saturate QCR bound using single copy measurements. We may still need in general to have many repetitions of the experiment (many copies of a quantum state) to really be sure that the estimator that asymptotically saturates the CR bound (e.g. max-likelihood estimator) exists.

6.2 Multi-parameter case

Let us now consider a multi-parameter estimation problem, where the family of states ρ_θ is parametrized by K real parameters $\theta = \{\theta_1, \dots, \theta_K\}$. Similarly as in the classical case the following multiparameter generalization of the CR bound holds.

Theorem 6.2 (Multiparameter quantum Cramér-Rao bound). Given a family of states $\rho_{\boldsymbol{\theta}}$, $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_K\}$, the following matrix inequality holds:

$$\mathbb{C} \geq \mathbb{F}_Q^{-1}, \quad (6.32)$$

where \mathbb{C} is the $K \times K$ covariance matrix corresponding to estimation involving any locally unbiased estimators and arbitrary measurements and \mathbb{F}_Q is the QFI matrix defined as:

$$(\mathbb{F}_Q)_{ij} = \frac{1}{2} \text{Tr} [\rho_{\boldsymbol{\theta}} (\Lambda_{\boldsymbol{\theta},i} \Lambda_{\boldsymbol{\theta},j} + \Lambda_{\boldsymbol{\theta},j} \Lambda_{\boldsymbol{\theta},i})], \quad (6.33)$$

where $\Lambda_{\boldsymbol{\theta},i}$ is the SLD corresponding to parameter θ_i :

$$\frac{d\rho_{\boldsymbol{\theta}}}{d\theta_i} = \frac{1}{2} (\rho_{\boldsymbol{\theta}} \Lambda_{\boldsymbol{\theta},i} + \Lambda_{\boldsymbol{\theta},i} \rho_{\boldsymbol{\theta}}). \quad (6.34)$$

Proof. The proof utilizes the same steps that could be found in earlier derivations of the single parameter quantum CR bound and multiparameter classical CR bound. We provide the proof below, without comments as we basically repeat the steps that were employed in the earlier proofs:

$$\begin{aligned} v^T \mathbb{F}_Q v &= \int dx \frac{[\sum_i v_i \text{ReTr}(\Pi_x \Lambda_{\boldsymbol{\theta},i} \rho_{\boldsymbol{\theta}})]^2}{\text{Tr} \rho_{\boldsymbol{\theta}} \Pi_x} \leq \\ &= \int dx \frac{|\text{Tr}(\sum_i v_i \Pi_x \Lambda_{\boldsymbol{\theta},i} \rho_{\boldsymbol{\theta}})|^2}{\text{Tr} \rho_{\boldsymbol{\theta}} \Pi_x} \leq \\ &= \int dx \frac{\text{Tr}(\rho_{\boldsymbol{\theta}} \Pi_x) \text{Tr}(\sum_{ij} v_i v_j \sqrt{\rho_{\boldsymbol{\theta}}} \Lambda_{\boldsymbol{\theta},i} \sqrt{\Pi_x} \sqrt{\Pi_x} \Lambda_{\boldsymbol{\theta},j} \sqrt{\rho_{\boldsymbol{\theta}}})}{\text{Tr}(\rho_{\boldsymbol{\theta}} \Pi_x)} \underline{d}_B \\ &= \sum_{ij} v_i \text{Tr}(\rho_{\boldsymbol{\theta}} \Lambda_{\boldsymbol{\theta},i} \Lambda_{\boldsymbol{\theta},j}) v_j = \\ &= v^T \mathbb{F}_Q v. \quad (6.35) \end{aligned}$$

■

Remark. In the multiparameter case the QCR is not in general saturable. This is due to the fact that different SLDs corresponding to different parameters might not commute, so it is not clear whether there exist a single measurement that provides the optimal FI for all the parameters simultaneously. In fact, tighter bounds exist that are more informative and take into account the necessary trade-offs due to incompatibility of measurements which are optimal for different parameters, see

Sec. 6.4.

Remark. In order to meaningfully speak about the optimal multi-parameter estimation scenario it is natural to define a scalar cost function via $C = \text{Tr} \mathbb{C} G$, where G is some positive semidefinite cost-matrix, which encodes the relative weights we ascribe to uncertainties of different parameters (or combination of parameters) that we include in the final cost function. Note that the QCR bound implies $\text{Tr} \mathbb{C} G \geq \text{Tr} \mathbb{F}_Q G$.

6.3 Natural metric in the space of quantum states

Bures metric Since the QFI is a measure of distinguishability of quantum states we may employ it as a natural measure of distance between quantum states. Let us define infinitesimal distance between states ρ and $\rho + d\rho$ as:

$$d_B(\rho, \rho + d\rho)^2 = \frac{1}{4} \text{Tr}(\rho d\Lambda^2), \quad (6.36)$$

where $d\Lambda$ is defined via:

$$d\rho = \frac{1}{2} (d\Lambda \rho + \rho d\Lambda). \quad (6.37)$$

\underline{d}_B is called the Bures distance and the resulting metric in the space of quantum states is called the Bures metric. When restricted to pure states, the Bures metric is referred to as the Fubini-Study metric.

Fidelity In quantum information theory a commonly used measure of similarity of two quantum states is the so called *fidelity*. Given two pure states $|\psi_1\rangle, |\psi_2\rangle$ the fidelity is defined as:

$$\mathcal{F}(|\psi_1\rangle, |\psi_2\rangle) = |\langle \psi_1 | \psi_2 \rangle|^2 \quad (6.38)$$

and may be interpreted as the probability of observing state $|\psi_1\rangle$ as state $|\psi_2\rangle$ or vice versa.

Fidelity is generalized to mixed states using

the following formula:

$$\mathcal{F}(\rho_1, \rho_2) = \max_{|\psi_1\rangle, |\psi_2\rangle} |\langle \psi_1 | \psi_2 \rangle|^2, \quad (6.39)$$

where $|\psi_i\rangle \in \mathcal{H} \otimes \mathcal{H}_E$ are purifications of states ρ_i : $\rho_i = \text{Tr}_E |\psi_i\rangle \langle \psi_i|$, where E represents ancillary Hilbert space used for purification.

Theorem 6.3 (Uhlman theorem). An explicit formula for the fidelity between two mixed states reads:

$$\mathcal{F}(\rho_1, \rho_2) = (\text{Tr} |\sqrt{\rho_1} \sqrt{\rho_2}|)^2, \quad (6.40)$$

where $|A| = \sqrt{A^\dagger A}$.

Proof. Let $|\psi_i\rangle \in \mathcal{H}_S \otimes \mathcal{H}_E$ be purifications of ρ_i . We can rewrite each $|\psi_i\rangle$ as a matrix A_i , such that $(A_i)_l^k = \psi_{i,kl}$, where k, l represent indices corresponding to spaces \mathcal{H} and \mathcal{H}_E . Then $\rho_i = \text{Tr}_E |\psi_i\rangle \langle \psi_i| = A_i A_i^\dagger$. Now,

$$\mathcal{F}(\rho_1, \rho_2) = \max_{A_1, A_2} |\text{Tr}(A_1 A_2^\dagger)|^2, \quad \rho_i = A_i A_i^\dagger. \quad (6.41)$$

Note that changing a purification for a given ρ corresponds to replacing $A \rightarrow AU$, where U is a unitary. Let us now consider the polar decomposition of A_i , $A_i = \sqrt{\rho_i} U_i$, and observe that

$$\begin{aligned} |\text{Tr}(A_1 A_2^\dagger)| &= |\text{Tr}(\sqrt{\rho_1} U_1 U_2^\dagger \sqrt{\rho_2})| = \\ &|\text{Tr}(\sqrt{\rho_2} \sqrt{\rho_1} U_1 U_2^\dagger)| \leq \text{Tr} |\sqrt{\rho_1} \sqrt{\rho_2}|. \end{aligned} \quad (6.42)$$

Note that the above inequality can be saturated if we choose purifications such that $U_1 U_2^\dagger = \mathbb{1}$.

In the above derivation we have used the property that for any hermitian matrix A and unitary U , $|\text{Tr}(AU)| \leq \text{Tr}|A|$. This can be seen as follows. Let $A = \sum_i a_i |i\rangle \langle i|$ be eigendecomposition of A . Since trace is basis independent we can perform it using the basis $|i\rangle$: $|\text{Tr}(AU)| = |\sum_i a_i \langle i|U|i\rangle|$. Absolute value of any matrix element of a unitary matrix is smaller or equal to 1. Hence $|\sum_i a_i \langle i|U|i\rangle| \leq |\sum_i a_i| \leq \sum_i |a_i| = \text{Tr}|A|$. ■

Relation between the Fidelity and the QFI We now prove a theorem that provides a link between the fidelity and the QFI, by showing that infinitesimal change in the fidelity when a quantum state is changed is proportional to the QFI.

Theorem 6.4. For two infinitesimally close states $\rho_\theta, \rho_{\theta+d\theta}$,

$$\mathcal{F}(\rho_\theta, \rho_{\theta+d\theta}) = 1 - \frac{1}{4} F_Q(\rho_\theta) d\theta^2 + O(d\theta^3). \quad (6.43)$$

Proof. Consider:

$$\begin{aligned} \sqrt{\mathcal{F}(\rho_\theta, \rho_{\theta+d\theta})} &= \text{Tr} [\sqrt{\rho_\theta} (\rho_\theta + \dot{\rho}_\theta d\theta) \sqrt{\rho_\theta}] = \\ &= \text{Tr} [\rho_\theta^2 + \sqrt{\rho_\theta} \dot{\rho}_\theta \sqrt{\rho_\theta} d\theta], \end{aligned} \quad (6.44)$$

where we have made a replacement $\rho_{\theta+d\theta} = \rho_\theta + \dot{\rho}_\theta d\theta$. We now want to expand the above quantity up to the second order in $d\theta$. Since we deal with operators we have to be careful. We write:

$$\text{Tr} [\rho_\theta^2 + \sqrt{\rho_\theta} \dot{\rho}_\theta \sqrt{\rho_\theta} d\theta] = \rho_\theta + Ad\theta + Bd\theta^2 + O(d\theta^3), \quad (6.45)$$

where A and B are operators we want to determine now. Let us take square of the both sides of the above equations and compare terms in the leading orders in $d\theta$. As a result we obtain the following equations:

$$\rho_\theta A + A \rho_\theta = \sqrt{\rho_\theta} \dot{\rho}_\theta \sqrt{\rho_\theta} \quad (6.46)$$

$$A^2 + \rho_\theta B + B \rho_\theta = 0. \quad (6.47)$$

Solving the above equations in the eigenbasis of $\rho_\theta = \sum_i p_i |i\rangle \langle i|$ we get:

$$A_{ij} = \frac{\sqrt{p_i p_j}}{p_i + p_j} \dot{\rho}_{\theta,ij} \rightarrow \text{Tr}(A) = 0, \quad (6.48)$$

while

$$\begin{aligned} B_{ij} &= -\frac{\sum_k A_{ik} A_{kj}}{p_i + p_j} = \\ &= -\frac{1}{p_i + p_j} \sum_k \frac{\sqrt{p_i p_j}}{p_i + p_k} \frac{\sqrt{p_k p_j}}{p_k + p_j} \dot{\rho}_{\theta,ik} \dot{\rho}_{\theta,kj}. \end{aligned} \quad (6.49)$$

As a result

$$\text{Tr} B = -\sum_{ik} \frac{p_k}{2(p_i + p_k)} |\dot{\rho}_{\theta,ik}|^2. \quad (6.50)$$

Note that from the definition of the SLD, $\dot{\rho}_{\theta,ik} = \frac{1}{2} \Lambda_{\theta,ik} (p_i + p_k)$, and hence $\text{Tr} B = -\frac{1}{2} \text{Tr} \rho_\theta \Lambda_\theta^2 = -\frac{1}{8} F_Q$. So we get

$$\sqrt{\mathcal{F}(\rho_\theta, \rho_{\theta+d\theta})} = 1 - \frac{1}{8} F_Q d\theta^2 + O(d\theta^3) \quad (6.51)$$

which yields the desired theorem when squared. ■

Remark. Combining the claims of Theorems 6.3, ?? it immediately follows that

$$F_Q(\rho_\theta) = \min_{|\psi_\theta\rangle} F_Q(|\psi_\theta\rangle), \quad (6.52)$$

where minimization is performed over all pu-

rifications of the mixed states ρ_θ .

6.4 Holevo Cramer-Rao bound

Below we present Holevo Cramer-Rao bound for multiparameter estimation, which is tighter than the QFI one, but unfortunately also harder to be applied, since we will not be able to completely get rid of the problem of optimization over the measurements.

Theorem 6.5 (Holevo Cramér-Rao bound). Given a family of states ρ_θ , $\theta = \{\theta_1, \dots, \theta_K\}$, then for any locally unbiased measurement/estimation strategy and any cost matrix G the following bound on the estimation cost holds:

$$\text{Tr}(G \cdot \mathbb{C}) \geq \min_{X_i \in \mathcal{L}(\mathcal{H})} \text{Tr}(G \cdot \text{Re}V) + \|\sqrt{G} \cdot \text{Im}V \sqrt{G}\|_1, \quad (\mathbb{C} - \text{Re}V) \geq \sum_l |\lambda_l| |l\rangle \langle l| = \text{Abs}(\text{Im}V). \quad (6.60)$$

where $\|\cdot\|_1$ is the trace norm ($\|A\|_1 := \text{Tr}\sqrt{AA^\dagger}$), $V_{ij} := \text{Tr}(\rho_\theta X_i X_j)$ and X_j are hermitian matrices satisfying $\text{Tr}(\frac{d\rho_\theta}{d\theta_i} X_j) = \delta_{ij}$.

Proof. For any measurement Π_x and estimator $\tilde{\theta}(x)$ we may define

$$X_i := \sum_x (\tilde{\theta}_i(x) - \theta_i) \Pi_x. \quad (6.54)$$

Therefore \min_{X_i} in (6.53) corresponds to optimization over measurements. Since we work in the frequentist approach, we as usual impose the local unbiasedness condition:

$$\frac{d}{d\theta_j} \sum_x \text{Tr}(\rho_\theta \Pi_x) \tilde{\theta}_i(x) = \delta_{ij}, \quad (6.55)$$

which when written in terms of X_i implies

$$\text{Tr}\left(\frac{d\rho_\theta}{d\theta_i} X_j\right) = \delta_{ij}. \quad (6.56)$$

For simplicity of notation let us focus on estimation around $\theta = 0$. First, we proof matrix inequality $\mathbb{C} \geq V$ (for any set of X_i). To do that we introduce following matrix, acting on $\mathbb{C}^K \otimes \mathcal{H}$ (where \mathbb{C}^K is an abstract K -dimensional Hilbert space, where K is the number

of parameters in the problem):

$$\sum_x \begin{bmatrix} \tilde{\theta}_1(x) - X_1 \\ \vdots \\ \tilde{\theta}_K(x) - X_K \end{bmatrix} \Pi_x [\dots, \tilde{\theta}_i(x) - X_i, \dots] \geq 0. \quad (6.57)$$

As any x element of this sum is positive-defined and hence is the whole matrix. After transferring all terms with X_i to the right side, substituting (6.54) in proper places, using $\sum_j \Pi_j = \mathbb{1}$ and applying $\text{Tr}_{\mathcal{H}}(\rho_\theta \cdot)$ we arrive at

$$\mathbb{C} \geq V. \quad (6.58)$$

Note, that this inequality holds on a complex space \mathbb{C}^K and, as it may happen that $\text{Im}V \neq 0$, the resulting inequality is stronger than simply $\text{Tr}(G \cdot \mathbb{C}) \geq \text{Tr}(G \cdot V)$ and stronger bounds in principle be derived.

Thanks to the fact that \mathbb{C} is a symmetric matrix, from $\mathbb{C} \geq V$ we also have $\mathbb{C} \geq V^T$. Taking into account that the V matrix is hermitian, we may therefore write:

$$\begin{aligned} (\mathbb{C} - \text{Re}V) &\geq i\text{Im}V, \\ (\mathbb{C} - \text{Re}V) &\geq -i\text{Im}V, \end{aligned} \quad (6.59)$$

Introducing the eigenbasis of $i\text{Im}V = \sum_l \lambda_l |l\rangle \langle l|$, the two above inequalities are equivalent to:

where $\text{Abs}(A)$ represents a matrix with the same eigenvectors as A but with all eigenvalues replaced with their absolute values. Analogously, since G is positive we may multiply both sides (6.59) by \sqrt{G} without affecting the validity of inequalities and arrive at:

$$\sqrt{G}(\mathbb{C} - \text{Re}V)\sqrt{G} \geq \text{Abs}(\sqrt{G}\text{Im}V\sqrt{G}) = |\sqrt{G}\text{Im}V\sqrt{G}|, \quad (6.61)$$

where $|A| = \sqrt{A^\dagger A}$ and we have used the fact that $\text{Abs}(A) = |A|$ for unitary diagonalizable matrixes (i.e. normal matrixes— here $\sqrt{G}\text{Im}V\sqrt{G}$ is anti-hermitian and hence normal). Finally tracing the above inequality we get

$$\text{Tr}(G \cdot \mathbb{C}) \geq \text{Tr}(G \cdot \text{Re}V) + \|\sqrt{G} \cdot \text{Im}V \sqrt{G}\|_1. \quad (6.62)$$

The above inequality holds for any measurement (which defines X_i matrices simultaneously) and we may tighten it by choosing a set X_i which the minimizes the right-hand side for given cost function G (without bothering if there exists corresponding measurement). In this way we arrive at (6.53). ■

Note, that for projective measurements (i.e. $\Pi_x \Pi_{x'} = \delta_{x,x'} \Pi_x$) inequality $\mathbb{C} \geq V$ becomes equality $\mathbb{C} = V$ and that automatically implies $\text{Im}V = 0$. As any POVM on the \mathcal{H} may

be modeled as a projective measurement on $\mathcal{H} \oplus \mathcal{H}_M$, another version of above theorem may be formulated:

Theorem 6.6 (Matsumoto Cramér-Rao bound). Given a family of states ρ_θ , $\theta = \{\theta_1, \dots, \theta_K\}$, then for any locally unbiased measurement/estimation strategy and any cost matrix G the following bound on the estimation cost holds:

$$\mathrm{Tr}(G \cdot \mathbb{C}) \geq \min_{X_i \in \mathcal{L}(\mathcal{H} \oplus \mathbb{C}^K): \mathrm{Im}V=0} \mathrm{Tr}(G \cdot V), \quad (6.63)$$

where $V_{ij} := \mathrm{Tr}(\rho_\theta X_i X_j)$ and X_j are hermitian matrices and satisfy $\mathrm{Tr}(\frac{d\rho_\theta}{d\theta_i} X_j) = \delta_{ij}$. This bound is exactly equivalent to the Holevo Cramér-Rao bound

Proof. It is enough to prove the equivalence with the Holevo Cramér-Rao bound. Let us decompose $X_i \in \mathcal{L}(\mathcal{H} \oplus \mathbb{C}^K)$ into $X_i = Y_i + Z_i$, where $P_{\mathcal{H}} Y P_{\mathcal{H}} = Y$ and $P_{\mathcal{H}} Z P_{\mathcal{H}} = 0$ ($P_{\mathcal{H}}$ is projection onto \mathcal{H} here). We define matrices V_X, V_Y, V_Z as $(V_X)_{ij} = \mathrm{Tr}(\rho_\theta X_i X_j)$ etc. To prove equivalence between the Matsumoto and the Holevo Cramér-Rao bound, we show that for any fixed V_Y the following equation holds:

$$\min_{Z_i: \mathrm{Im}V_X=0} \mathrm{Tr}(G \cdot V_X) = \mathrm{Tr}(G \cdot \mathrm{Re}V_Y) + \|G \cdot \mathrm{Im}V_Y\|_1. \quad (6.64)$$

Directly from the definition of Y_i, Z_i we have $V_X = V_Y + V_Z$. After substituting it to the above equation the whole problem simplifies to:

$$\min_{Z_i: \mathrm{Im}V_Z = -\mathrm{Im}V_Y} \mathrm{Tr}(G \cdot \mathrm{Re}V_Z) = \|G \cdot \mathrm{Im}V_Y\|_1. \quad (6.65)$$

First, since $\sqrt{G} V_Z \sqrt{G} \geq 0$ we have $\mathrm{Tr}(G \cdot \mathrm{Re}V_Z) \geq \|G \cdot \mathrm{Im}V_Z\|_1 = \|G \cdot \mathrm{Im}V_Y\|_1$. The only thing left to do is to show that this inequality is saturable.

Let us take $V_Z = \mathrm{Abs}(\mathrm{Im}V_Y) - i\mathrm{Im}V_Y$. Such a V_Z is positive, saturates the above inequality and as a result there exists a corresponding set of Z_i for which the equality (6.64) holds. ■

Remark. To compare the Holevo CR bound with the standard QFI based CR bound, let us come back to the weaker bound which appeared in the first proof, $\mathrm{Tr}(G \cdot \mathbb{C}) \geq \mathrm{Tr}(G \cdot V)$. We show below by using the Lagrange multiplier method, that

$$\min_{X_i} \mathrm{Tr}(G \cdot V) = \mathrm{Tr}(G \cdot \mathbb{F}_Q^{-1}). \quad (6.66)$$

and hence this weaker inequality corresponds to the standard multi-parameter CR bound.

Proof. Keeping in mind the local unbiasedness constraint:

$$\mathrm{Tr}\left(\frac{d\rho_\theta}{d\theta_i} X_j\right) = \frac{1}{2} \mathrm{Tr}(\rho_\varphi \{X_i, L_j\}) = \delta_{ij}. \quad (6.67)$$

let us write the solution to the minimization problem of the r.h.s of Eq. (6.66) explicitly using the Lagrange multiplier method. Introducing Lagrange multipliers λ_{ij} we need to minimize

$$\frac{1}{2} \sum_{ij} G_{ij} \mathrm{Tr}(\rho_\varphi \{X_i, X_j\}) - \lambda_{ij} [\delta_{ij} - \frac{1}{2} \mathrm{Tr}(\rho_\varphi \{X_i, L_j\})] \quad (6.68)$$

over Hermitian X_i . Each n -dimensional Hermitian matrix X_i may be parametrized by n^2 real parameters. Taking the derivatives over each of these produces a set of matrix equations,

$$\forall_i \sum_j G_{ij} \{\rho_\varphi, X_j\} - \lambda_{ij} \{\rho_\varphi, L_j\} = 0. \quad (6.69)$$

Taking

$$X_i = \sum_j (G^{-1} \Lambda)_{ij} L_j. \quad (6.70)$$

where by Λ we denote the matrix of Lagrange multipliers $(\Lambda)_{ij} = \lambda_{ij}$ it is clear that Eq. (6.69) is satisfied. Moreover, the constraint condition $\frac{1}{2} \mathrm{Tr}(\{X_i, L_j\} \rho_\varphi) = \delta_{ij}$ reads:

$$\frac{1}{2} (G^{-1} \Lambda)_{ik} \mathrm{Tr}(\{L_k, L_j\} \rho_\varphi) = \delta_{ij}. \quad (6.71)$$

This implies that the Lagrange multiplier matrix must be chosen so that:

$$G^{-1} \Lambda F_Q = \mathbb{1}. \quad (6.72)$$

As a result the solution to the minimization problem reads

$$X_i = \sum_j (F_Q^{-1})_{ij} L_j \quad (6.73)$$

and utilizing the fact that QFI matrix is symmetric we get

$$\mathrm{Tr}(G \cdot \mathrm{Re}V) = \mathrm{Tr}(G F_Q^{-1} F_Q F_Q^{-1}) = \mathrm{Tr}(G F_Q^{-1}), \quad (6.74)$$

which ends the proof. ■

Even though the Holevo CR bound is tighter than the QFI one, there is still no guarantee for saturability. Note, that while any measurement with estimator define X_i , there is no guarantee that for any set of X_i there exist proper Π_x and $\tilde{\theta}(x)$ satisfying (6.54). However, it may be proved that they exist when

we consider pure states: $\rho_{\theta} = |\psi_{\theta}\rangle\langle\psi_{\theta}|$.

Theorem 6.7 (Saturability of the Holevo and Matsumoto Cramér-Rao bound for pure states). Given a family of pure states $\rho_{\theta} = |\psi_{\theta}\rangle\langle\psi_{\theta}|$, $\theta = \{\theta_1, \dots, \theta_K\}$ and demanding the measurement/estimation strategy to be locally unbiased, we have that for any cost matrix G :

$$\begin{aligned} \min_{\Pi_x, \tilde{\theta}(x)} \text{Tr}(G \cdot \mathbb{C}) &= \\ &= \min_{X_i \in \mathcal{L}(\mathcal{H} \oplus \mathbb{C}^K): \text{Im}V=0} \text{Tr}(G \cdot V) \\ &= \min_{X_i \in \mathcal{L}(\mathcal{H})} \text{Tr}(G \cdot \text{Re}V) + \|G \cdot \text{Im}V\|_1, \end{aligned} \quad (6.75)$$

where $V_{ij} := \text{Tr}(\rho_{\theta} X_i X_j)$ and X_j satisfy $\text{Tr}(\frac{d\rho_{\theta}}{d\theta_i} X_j) = \delta_{ij}$.

Proof. We will prove the first equality (the second one comes directly from theorem 6.6).

As we deal only with pure states, let us introduce simplified notation: $|x_i\rangle := X_i |\psi_{\theta}\rangle$ (and then $V_{ij} = \langle x_i | x_j \rangle$). Note, that $|x_i\rangle$ which minimize the above expression, always satisfy $\forall_i \langle \psi_{\theta} | x_i \rangle = 0$ (as adding $|x'_i\rangle = |x_i\rangle + \alpha_i |\psi_{\theta}\rangle$ may only increase V_{ij} and it does not affect the constraint $\text{Tr}(\frac{d\rho_{\theta}}{d\theta_i} X_j) = \delta_{ij}$).

As $\forall_i \langle \psi_{\theta} | x_i \rangle = 0$ and $\forall_{ij} \langle x_i | x_j \rangle \in \mathbb{R}$ we may choose a basis $\{|b_i\rangle\}$ of $\text{span}\{|\psi_{\theta}\rangle, |x_1\rangle, \dots, |x_K\rangle\}$ satisfying: $\forall_i \langle \psi_{\theta} | b_i \rangle \in \mathbb{R}_{\setminus 0}$ and $\forall_{ij} \langle x_i | b_j \rangle \in \mathbb{R}$.

Then we can define a projective measurement:

$$\Pi_j = |b_j\rangle\langle b_j| \quad (j = 1, \dots, K+1), \quad \Pi_0 = \mathbb{1} - \sum_{j=1}^{K+1} |b_j\rangle\langle b_j| \quad (6.76)$$

with estimator:

$$\tilde{\theta}_i(j) = \frac{\langle b_j | x_i \rangle}{\langle b_j | \psi_{\theta} \rangle} + \theta_i \quad (j = 1, \dots, K+1), \quad \tilde{\theta}_i(0) = 0 \quad (6.77)$$

which is locally unbiased and satisfies

$$|x_i\rangle = \sum_{j=1}^{K+1} (\tilde{\theta}_i(j) - \theta_i) \Pi_j |\psi_{\theta}\rangle. \quad (6.78)$$

Therefore, by virtue of theorems 6.5 and 6.6, the thesis is proved. ■

In case of mixed states the bound is not always saturable. However, it is always saturable in the asymptotic limit of many copies

$\rho_{\theta}^{(N)} = \rho_{\theta}^{\otimes N}$ in the sense that:

$$\begin{aligned} \lim_{N \rightarrow \infty} N \min_{\Pi_x, \tilde{\theta}(x)} \text{Tr}(G \cdot \mathbb{C}) &= \\ \min_{X_i \in \mathcal{L}(\mathcal{H})} \text{Tr}(G \cdot \text{Re}V) + \|\sqrt{G} \cdot \text{Im}V \sqrt{G}\|_1. \end{aligned} \quad (6.79)$$

where in the above formula the left-hand side minimum reference to minimization over all measurement and estimation strategies on N -copy state $\rho_{\theta}^{(N)}$, while when calculating the Holevo CR bound on the right-hand side we take a single copy state ρ_{θ} . This is the consequence of the quantum local asymptotic normality theorem and the fact that Holevo CR bound is saturable for gaussian states models.

Problems

Problem 6.1 Consider the following family of qubit states:

$$|\psi_\varphi\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)\exp(i\varphi)|1\rangle, \quad (6.80)$$

where parameter θ we regard as known while our goal is to estimate φ .

- What does quantum Cramer-Rao bound tell us about achievable estimation precision if we were given N copies of the above state: $|\psi_\varphi\rangle^{\otimes N}$
- What would be the optimal measurement that guarantees saturation of the Cramer-Rao bound for large N —does the measurement depend on the value of estimated parameter φ ?
- Repeat the above points, in case where instead of N copies of the pure state $|\psi_\varphi\rangle$ we get N noisy copies each described by the following mixed state: $\rho_\varphi = p|\psi_\varphi\rangle\langle\psi_\varphi| + (1-p)\mathbb{1}/2$, where p is a known parameter.
- Consider multiparameter estimation case, where we assume that apart from φ also θ and p are unknown parameters to be estimated. Write down the QFI matrix and try to conjecture whether there is measurement that allows to saturation of the CR bound for all parameters simultaneously.

Problem 6.2 We have introduced the Bures metric, where the distance element is defined through:

$$d_B^2\rho = \frac{1}{4}\text{Tr}(\rho d\Lambda^2), \quad d\rho = \frac{1}{2}(d\Lambda\rho + \rho d\Lambda). \quad (6.81)$$

Prove that in case of a qubit, for which a general state is parameterized as:

$$\rho = \frac{1}{2}(\mathbb{1} + \vec{r} \cdot \vec{\sigma}), \quad \vec{r} = (r \sin \theta \cos \varphi, r \sin \theta \sin \varphi, r \cos \theta), \quad (6.82)$$

Bures metric takes the form:

$$d_B^2\rho = \frac{1}{4} \left[\frac{d^2 r}{1-r^2} + r^2(d^2\theta + \sin^2\theta d^2\varphi) \right]. \quad (6.83)$$

Hint: Start by parameterizing the state using Cartesian coordinates (x,y,z) and try to find Λ from condition (6.81)—remember that Λ is hermitian which means that it can be written as $\Lambda = \sum_{i=0}^3 \lambda_i \sigma_i$, where $\lambda_i \in \mathbb{R}$, and $\sigma_0 = \mathbb{1}$.

Problem 6.3 Some problem for Holevo.

Chapter 7

Quantum Bayesian estimation

In this chapter we will follow the Bayesian paradigm and develop methods to find optimal measurement and estimation strategies for quantum estimation models. A Bayesian quantum estimation model consists of the family of states ρ_θ and the prior distribution $p(\theta)$. The goal is to find a measurement $\{\Pi_x\}$ and an estimator $\tilde{\theta}(x)$ that minimize the average Bayesian cost:

$$\bar{C} = \int d\theta p(\theta) \int dx \text{Tr}(\rho_\theta \Pi_x) C[\theta, \tilde{\theta}(x)], \quad (7.1)$$

where $C(\theta, \tilde{\theta})$ is the cost function penalizing for the deviation of the estimator from the true value. In particular, if we chose $C(\theta, \tilde{\theta}) = (\theta - \tilde{\theta})^2$ we return to the standard Bayesian variance cost function.

From chapter 5 we now that once the measurement is fixed and hence we can write the conditional probability $p(x|\theta)$ it is clear how to find the optimal Bayesian estimator. Still, the issue of determining the optimal measurement remains non-trivial.

Note that we can formally relabel the measurement operators Π_x to $\Pi_{\tilde{\theta}(x)}$, so that in fact the label represents the estimated value of parameter, $\Pi_{\tilde{\theta}} = \int dx \Pi_x \delta(\tilde{\theta} - \tilde{\theta}(x))$. We do not lose any generality here, but thanks to this we can combine the double minimization over the measurement and the estimator to a single

optimization over the measurements only:

$$\min_{\{\Pi_{\tilde{\theta}}\}} \bar{C}, \quad \Pi_{\tilde{\theta}} \geq 0, \quad \int d\tilde{\theta} \Pi_{\tilde{\theta}} = \mathbb{1} \quad (7.2)$$

$$\bar{C} = \int d\theta d\tilde{\theta} p(\theta) \text{Tr}(\rho_\theta \Pi_{\tilde{\theta}}) C(\theta, \tilde{\theta}). \quad (7.3)$$

Of course this in general is a untractable problem, as the space of all allowed generalized measurements is enormous. Still, as demonstrated below with some additional assumptions on the cost function or the set of states, the problem may be solved. Note that the above reformulation makes the classical results on the optimal Bayesian estimation not really very helpful in deriving fundamental bounds on precision, as we have incorporated the estimator in the labeling of the measurement operators and hence in some sense the optimal estimator is given for free once we solve the above search for the optimal measurement. This is not to say, that we will never utilize classical results. Quite contrary, whenever we will desire to provide a practical protocol that performs optimally we will eventually be forced to write down a standard (projective) measurement and the explicit form of estimator that achieves the bound derived using formal approach formulated above, and then we will definitely make use of the optimal Bayesian estimator construction known from classical theory.

7.1 Quadratic cost problems

Let us start by restricting ourselves to the quadratic cost function $C(\theta, \tilde{\theta}) = (\theta - \tilde{\theta})^2$. For simplicity of the formulas that follow we redefine the parameter θ so that the expectation value of the prior distribution is zero, $\int d\theta p(\theta)\theta = 0$. The Bayesian variance to be minimized takes the form:

$$\begin{aligned} \overline{\Delta^2 \tilde{\theta}} &= \int d\theta d\tilde{\theta} p(\theta) \text{Tr}[\rho_\theta \Pi_{\tilde{\theta}} (\theta - \tilde{\theta})^2] = \\ &= \int d\theta p(\theta) \theta^2 + \text{Tr} \left[\int d\theta p(\theta) \rho_\theta \int d\tilde{\theta} \Pi_{\tilde{\theta}} \tilde{\theta}^2 \right] + \\ &\quad - 2 \text{Tr} \left[\int d\theta p(\theta) \theta \rho_\theta \int d\tilde{\theta} \Pi_{\tilde{\theta}} \tilde{\theta} \right] = \\ &= \Delta^2 \theta + \text{Tr}(\bar{\rho} \Lambda_2) - 2 \text{Tr}(\bar{\rho}' \Lambda_1), \end{aligned} \quad (7.4)$$

where we $\Delta^2 \theta = \int d\theta p(\theta) \theta^2$ represents the variance of the prior distribution, $\bar{\rho} = \int d\theta p(\theta) \rho_\theta$ is the average state, $\bar{\rho}' = \int d\theta p(\theta) \theta \rho_\theta$ and $\Lambda_k = \int d\tilde{\theta} \Pi_{\tilde{\theta}} \tilde{\theta}^k$. The following theorem determines the minimum of the above quantity optimized over all measurements $\Pi_{\tilde{\theta}}$.

Theorem 7.1. Given family of states ρ_θ and the priori distribution $p(\theta)$ (with expectation value at $\theta = 0$) the minimal Bayesian variance for estimation of θ is given by:

$$\overline{\Delta^2 \tilde{\theta}} = \Delta^2 \theta - \text{Tr}(\bar{\rho} \bar{\Lambda}^2), \quad (7.5)$$

where $\bar{\Lambda}$ is defined by the following equation:

$$\bar{\rho}' = \frac{1}{2} (\bar{\Lambda} \bar{\rho} + \bar{\rho} \bar{\Lambda}) \quad (7.6)$$

and $\bar{\rho} = \int d\theta p(\theta) \rho_\theta$, $\bar{\rho}' = \int d\theta p(\theta) \theta \rho_\theta$.

Proof. Let us first prove that if a given POVM measurement $\{\Pi_{\tilde{x}}\}$ is optimal, then we may find a projective measurement yielding the same cost. Let us perform eigen-decomposition of Λ_1 operator:

$$\Lambda_1 = \int d\tilde{\theta} \Pi_{\tilde{\theta}} \tilde{\theta} = \sum_i \tilde{\theta}_i |\tilde{\theta}_i\rangle \langle \tilde{\theta}_i|. \quad (7.7)$$

Consider now the following inequality:

$$\int d\tilde{\theta} (\tilde{\theta} - \Lambda_1) \Pi_{\tilde{\theta}} (\tilde{\theta} - \Lambda_1) \geq 0, \quad (7.8)$$

which is true since $\Pi_{\tilde{\theta}} \geq 0$ while Λ_1 is hermitian. This implies:

$$\int d\tilde{\theta} \Pi_{\tilde{\theta}} \tilde{\theta}^2 + \Lambda_1^2 - 2\Lambda_1 \geq 0 \quad (7.9)$$

and hence

$$\Lambda_2 \geq \Lambda_1^2. \quad (7.10)$$

Let us now replace the measurement $\{\Pi_{\tilde{\theta}}\}$ with the projective measurement, corresponding to the projection on the eigenbasis $|\tilde{\theta}_i\rangle$ of Λ_1 . For this choice $\Lambda_2 = \Lambda_1^2$, which according to (7.10) is the smallest operator possible. Inspecting (7.4) we see that we want the term $\text{Tr}(\bar{\rho} \Lambda_2)$ to be as small as possible, and hence it is always optimal to choose the projective measurement in the eigenbasis of Λ_1 .

Assuming the measurement is projective, we may now introduce a single operator variable write $\bar{\Lambda} = \Lambda_1$, $\Lambda_2 = \bar{\Lambda}^2$ and the optimization problem amounts to minimization of the following cost function over a single hermitian operator $\bar{\Lambda}$:

$$\overline{\Delta^2 \tilde{\theta}} = \Delta^2 \theta - \text{Tr}(\bar{\rho} \bar{\Lambda}^2) - 2 \text{Tr}(\bar{\rho}' \bar{\Lambda}). \quad (7.11)$$

Since the above formula is quadratic in matrix $\bar{\Lambda}$, the minimization can be performed explicitly and the condition for finishing first derivative amounts to the following linear equation:

$$\bar{\Lambda} \bar{\rho} + \bar{\rho} \bar{\Lambda} - 2\bar{\rho}' = 0. \quad (7.12)$$

Multiplying the above equality by $\bar{\Lambda}$ and taking the trace of both sides we get that $\text{Tr}(\bar{\rho}' \bar{\Lambda}) = \text{Tr}(\bar{\rho} \bar{\Lambda}^2)$ and therefore we arrive at the formula stated in the theorem. ■

Gaussian prior distribution. Equations (7.5,7.6) remind of formulas used for calculation of the QFI. The main difference is that instead of the derivative of the state on the left hand side of (7.6) there appear the $\bar{\rho}'$ operator. In order to establish a closer relation between these two approaches consider a gaussian prior distribution $p_{\theta_0} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\theta-\theta_0)^2}{2\sigma^2}}$, where θ_0 is a free parameter determining the center of the prior. We will consider the effect of variation of the center of the prior around

$\theta_0 = 0$. We now have:

$$\begin{aligned} \bar{\rho}'|_{\theta_0=0} &= \int d\theta p_{\theta_0=0}(\theta)\theta\rho_\theta = \\ & \int d\theta \left. \frac{dp_{\theta_0}(\theta)}{d\theta_0} \right|_{\theta_0=0} \rho_\theta \sigma^2 = \\ \frac{d}{d\theta_0} \int d\theta p_{\theta_0}(\theta)\rho_\theta \Big|_{\theta_0=0} &= \sigma^2 \left. \frac{d\bar{\rho}_{\theta_0}}{d\theta_0} \right|_{\theta_0=0} = \\ & \Delta^2\theta \left. \frac{d\bar{\rho}_{\theta_0}}{d\theta_0} \right|_{\theta_0=0}. \end{aligned} \quad (7.13)$$

From the above formula we see that for Gaussian prior $\bar{\rho}'$ is proportional to the derivative of the averaged state $\bar{\rho}$ with respect to the shift of the prior distribution and the proportionality constant is the prior variance. Therefore:

$$\frac{1}{2} (\bar{\rho}\bar{\Lambda} + \bar{\Lambda}\bar{\rho}) = \Delta^2\theta \frac{d\bar{\rho}_{\theta_0}}{d\theta_0} \quad (7.14)$$

and consequently $\bar{\Lambda} = \Lambda\Delta^2\theta$, where Λ is the standard SLD for the QFI estimation approach where using $\bar{\rho}_{\theta_0}$ we want to estimate changes in the center of the prior θ_0 around point $\theta_0 = 0$. As a result: $\text{Tr}(\bar{\rho}\bar{\Lambda}^2) = (\Delta^2\theta)^2 F_Q(\bar{\rho}_{\theta_0})$. Finally we can write:

$$\overline{\Delta^2\bar{\theta}} = \Delta^2\theta [1 - \Delta^2\theta F_Q(\bar{\rho}_{\theta_0})], \quad (7.15)$$

and we have arrived at the relation between the cost in the quantum Bayesian estimation and the QFI of the corresponding problem of estimating the prior from the averaged state $\bar{\rho}_{\theta_0}$.

7.2 Covariant estimation problems

Even though the previous section has provided us with a general recipe how to find the optimal Bayesian strategy for quadratic cost functions this is not always enough in practical problems. Note, that unlike in the frequentist approach where we in fact work in the paradigm of small fluctuations of parameter value around a known value, in the Bayesian

approach we are typically faced with priors which are by no mean narrow and hence the approximation of arbitrary cost function to be locally quadratic is no longer justified.

This is especially pronounced in quantum estimation theory where we face problems where angle-like parameters, or more generally rotations are to be estimated. In such cases theory restricted to quadratic cost function is not really helpful. Unfortunately, if in the consideration from the previous chapter we replace the quadratic cost with some other cost function arbitrary one in general will not be able to provide a closed solution to the problem of determining the optimal measurement and hence the minimal cost.

Fortunately, as we show below if the problem enjoys certain symmetry, we may utilize some powerful methods based on group theory considerations and attack the problem of determining the optimal Bayesian estimation strategy from another perspective. We will refer to these class of estimation problems as covariant with respect to representation of a certain group.

Definition 7.1 (Covariant estimation problem). Let G be a Lie group where the group element $g \in G$ is the estimation parameter in our problem. Let U_g be a unitary representation of the group in some Hilbert space, $U_{g_1}U_{g_2} = U_{g_1g_2}$. We say that the Bayesian estimation problem is covariant with respect to U_g if and only if the following conditions are satisfied:

- a) The parameter to be estimated is an element of the group $g \in G$
- b) The family of states is generated by the action of the group representation $\rho_g = U_g\rho_eU_g^\dagger$ —is the orbit of the group.
- c) The cost function is left invariant with respect to the action of the group: $\forall_{g_1,g_2,h \in G} C(hg_1, hg_2) = C(g_1, g_2)$.
- d) The prior distribution is invariant with respect to the action of the group:

$dg p(g) = d(hg)p(hg)$ —the prior is uniform with respect to the Haar measure on the group.

Even though the above conditions appear quite restrictive they are nevertheless satisfied in a number of important quantum estimation problem, and in particular the phase estimation example we have been studying in the previous chapter.

Example 7.1 Consider the estimation problem, where the family of states

$$\rho_\varphi = |\psi_\varphi\rangle\langle\psi_\varphi|, \quad |\psi_\varphi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + e^{i\varphi}|1\rangle) \quad (7.16)$$

and we want to estimate φ given flat prior distribution $p(\varphi) = \frac{1}{2\pi}$. Note that the flat prior is a natural least informative choice for such a problem, and hence the issue of choosing the right prior that haunts the Bayesian approach is not relevant here. In order to formulate the complete model we need to choose a cost function. Note that the standard squared distance cost function is not a sensible choice as we are not working in the local phase estimation approach and hence the differences between the estimated and the real phase may be significant. Therefore, we want to use a cost function that takes into account the periodic nature of the phase parameter and does not penalizes us if the phase difference is a multiple of 2π . One natural choice is $C(\varphi, \tilde{\varphi}) = 4 \sin^2\left(\frac{\varphi - \tilde{\varphi}}{2}\right)$ —it reduces to the standard squared error for small phase deviations, takes maximum value for $\varphi - \tilde{\varphi} = \pi$ and reflects the periodic nature of the phase parameter.

This problem is indeed an example of a covariant estimation problem, where the group behind is the $U(1)$ group, with the representation $U_\varphi = e^{i\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\varphi}$ in the qubit space. The family of states forms indeed an orbit, where $|\psi_\varphi\rangle = U_\varphi|\psi_0\rangle$, $|\psi_0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Finally, the prior as well as the cost function are invariant under the action of the group, as it amounts to simple phase addition.

For covariant estimation problems the Bayesian cost is given by:

$$\bar{C} = \int dg d\tilde{g} \text{Tr}(U_g \rho_e U_g^\dagger \Pi_{\tilde{g}}) C(g, \tilde{g}), \quad (7.17)$$

where we assume that dg is the normalized Haar measure on the group, $\int dg = 1$, with respect to which the prior is trivial $p(g) = 1$.

Definition 7.2 (Covariant measurement). $\{\Pi_{\tilde{g}}\}$ is called a covariant measurement with

respect to an action of group representation U_g if and only if

$$\forall_{\tilde{g}, h} U_h \Pi_{\tilde{g}} U_h^\dagger = \Pi_{h\tilde{g}}. \quad (7.18)$$

Remark. In particular for a covariant measurement

$$\Pi_{\tilde{g}} = U_{\tilde{g}} \Pi_e U_{\tilde{g}}^\dagger, \quad (7.19)$$

so that all measurement operators are determined by a single *seed* operator Π_e .

Theorem 7.2 (Optimality of covariant measurements). For the covariant estimation problem, the optimal measurement can be found within the class of covariant measurements.

Proof. Let $\Pi_{\tilde{g}}^{\text{opt}}$ be the optimal measurement minimizing \bar{C} :

$$\bar{C}^{\text{opt}} = \int dg d\tilde{g} \text{Tr}(\Pi_{\tilde{g}}^{\text{opt}} \rho_g) C(g, \tilde{g}). \quad (7.20)$$

Let us define

$$\Pi_{\tilde{g}}^{\text{cov}} = \int dg' U_{g'}^\dagger \Pi_{g'\tilde{g}}^{\text{opt}} U_{g'}. \quad (7.21)$$

This is indeed a covariant measurement, since:

$$\begin{aligned} U_h \Pi_{\tilde{g}}^{\text{cov}} U_h^\dagger &= \int dg' U_{hg'^{-1}} \Pi_{g'\tilde{g}}^{\text{opt}} U_{g'h^{-1}} = \\ &\stackrel{g' \rightarrow g'h}{=} \int dg' U_{g'}^\dagger \Pi_{g'h\tilde{g}}^{\text{opt}} U_{g'} = \Pi_{h\tilde{g}}^{\text{cov}}. \end{aligned} \quad (7.22)$$

Moreover, the corresponding cost:

$$\begin{aligned} \bar{C}^{\text{cov}} &= \int dg d\tilde{g} \text{Tr}(\Pi_{\tilde{g}}^{\text{cov}} \rho_g) C(g, \tilde{g}) = \\ &\int dg d\tilde{g} \text{Tr} \left(\int dg' U_{g'}^\dagger \Pi_{g'\tilde{g}}^{\text{opt}} U_{g'} \rho_g U_{g'}^\dagger \right) C(g, \tilde{g}) = \\ &\int dg d\tilde{g} dg' \text{Tr} \left(U_{g'g}^\dagger \Pi_{g'\tilde{g}}^{\text{opt}} U_{g'g} \rho_g \right) C(g, \tilde{g}) = \\ &\stackrel{g \rightarrow g'g^{-1}}{\tilde{g} \rightarrow g'^{-1}\tilde{g}} \int dg d\tilde{g} dg' \text{Tr} \left(U_{\tilde{g}}^\dagger \Pi_{\tilde{g}}^{\text{opt}} U_{\tilde{g}} \rho_e \right) C(g'^{-1}g, g'^{-1}\tilde{g}) = \\ &\int dg d\tilde{g} dg' \text{Tr}(\Pi_{\tilde{g}}^{\text{opt}} \rho_g) C(g, \tilde{g}) = \bar{C}^{\text{opt}}. \end{aligned} \quad (7.23)$$

So is equal to the optimal minimal cost. ■

Thanks to the above theorem, the problem of identifying the optimal estimation strategy may be significantly simplified. Note that thanks to the covariance property of the mea-

surement we have:

$$\begin{aligned}\bar{C} &= \int dg d\tilde{g} \operatorname{Tr}(\Pi_{\tilde{g}}\rho_g) C(g, \tilde{g}) = \\ &= \int dg d\tilde{g} \operatorname{Tr}\left(U_{\tilde{g}^{-1}g}^\dagger \Pi_e U_{\tilde{g}^{-1}g} \rho_e\right) C(g, \tilde{g}) = \\ &\stackrel{g \rightarrow \tilde{g}g}{=} \int dg d\tilde{g} \operatorname{Tr}\left(U_g^\dagger \Pi_e U_g \rho_e\right) C(g, e) = \\ &= \int dg \operatorname{Tr}(\Pi_e \rho_g) C(g, e). \quad (7.24)\end{aligned}$$

The whole problem now amounts to minimization of the above quantity over a *single* operator Π_e with constraints $\Pi_e \geq 0$, $\int dg U_g \Pi_e U_g^\dagger = \mathbb{1}$. This is a huge simplification of the original problem and often the optimal operator Π_e may be found analytically, as is demonstrated in the examples that follow.

Example 7.2 (Phase estimation on product qubit states) Let us consider the problem of estimating the phase φ given N qubits:

$$\rho_\varphi^{(N)} = U_\varphi^{\otimes N} |\psi_0\rangle\langle\psi_0|^{\otimes N} U_\varphi^{\dagger \otimes N}, \quad |\psi_0\rangle = (|0\rangle + |1\rangle)/\sqrt{2}, \quad (7.25)$$

where $U_\varphi = e^{i \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \varphi}$. We assume the flat prior $p(\varphi) = 1/2\pi$, and the cost function $C(\varphi, \tilde{\varphi}) = 4 \sin^2(\frac{\varphi - \tilde{\varphi}}{2})$. The goal is to maximize

$$\bar{C}^{(N)} = \frac{2}{\pi} \int_0^{2\pi} d\varphi \operatorname{Tr}\left(\Pi_0^{(N)} \rho_\varphi^{(N)}\right) \sin^2 \varphi \quad (7.26)$$

over $\Pi_0^{(N)}$, such that $\Pi_0^{(N)} \geq 0$, $\int \frac{d\varphi}{2\pi} U_\varphi^{\otimes N} \Pi_0^{(N)} U_\varphi^{\dagger \otimes N} = \mathbb{1}$.

Let us first solve the $N = 1$ case:

$$\bar{C}^{(1)} = \frac{1}{\pi} \operatorname{Tr}\left(\Pi_0^{(1)} \int_0^{2\pi} d\varphi \sin^2 \varphi \begin{bmatrix} 1 & e^{-i\varphi} \\ e^{i\varphi} & 1 \end{bmatrix}\right). \quad (7.27)$$

Taking into account that $\sin^2 \varphi = \frac{1}{2} - \frac{1}{4}e^{i\varphi} - \frac{1}{4}e^{-i\varphi}$ and performing the integral in the above formula yields:

$$\bar{C}^{(1)} = \operatorname{Tr}\left(\Pi_0^{(1)} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}\right). \quad (7.28)$$

We need to minimize the above quantity over $\Pi_0^{(1)}$, keeping in mind that $\Pi_0^{(1)} \geq 0$, $\int \frac{d\varphi}{2\pi} U_\varphi \Pi_0^{(1)} U_\varphi^\dagger = \mathbb{1}$. The completeness condition imply that:

$$\Pi_0^{(1)} = \begin{bmatrix} 1 & a \\ a^* & 1 \end{bmatrix} \quad (7.29)$$

where a is an arbitrary complex number. However, the positivity condition further implies that $|a| \leq 1$. We

therefore need to perform the following minimization:

$$\begin{aligned}\min_{|a| \leq 1} \operatorname{Tr}\left(\begin{bmatrix} 1 & a \\ a^* & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}\right) &= \\ &= \min_{|a| \leq 1} (2 - \operatorname{Re}a) = 1 \quad (7.30)\end{aligned}$$

for $a = 1$. This means that the optimal seed measurement

$$\Pi_0^{(1)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = 2|+\rangle\langle+|. \quad (7.31)$$

Note that $\Pi_\varphi^{(1)} = 2|\psi_\varphi\rangle\langle\psi_\varphi|$ so the optimal measurement corresponds to POVM spanning all equatorial states. The resulting cost equals:

$$\bar{C}^{(1)} = 1. \quad (7.32)$$

One can check, that even though we have achieved the minimal cost using covariant measurement, the same cost would be obtained in this case using the simplest possible projective measurement with $\Pi_0 = |+\rangle\langle+|$, $\Pi_1 = |-\rangle\langle-|$ and the corresponding estimated values of the phase equal to $\tilde{\varphi}(0) = 0$, $\tilde{\varphi}(1) = \pi$.

We now move on to solve the general case $N > 1$. First of all note that $|\psi_\varphi\rangle^{\otimes N} \in \mathcal{H}_S^{\otimes N}$, where $\mathcal{H}_S^{\otimes N}$ is the fully symmetric (bosonic) subspace of the space of N qubits. Therefore without loosing generality we may restrict ourselves to this subspace. This space is $N + 1$ dimensional so it significantly reduces the size of the considered Hilbert space. Let us denote by $|n\rangle$ ($n = 0, \dots, N$) symmetric states with n qubits in states $|1\rangle$ and $N - n$ qubits in state 0 :

$$|n\rangle = \frac{1}{\sqrt{\binom{N}{n}}} \sum_{\text{perm}} |0, \dots, 0, \underbrace{1, \dots, 1}_n\rangle, \quad (7.33)$$

where the sum is performed over all non-trivial permutations. Then

$$|\psi_\varphi\rangle = \frac{1}{\sqrt{2^N}} \sqrt{\binom{N}{n}} e^{in\varphi} |n\rangle. \quad (7.34)$$

After the integration over φ is performed the formula for the cost to be minimized reduces to

$$\bar{C}^{(N)} = \frac{1}{2^{N-1}} \operatorname{Tr}\left(\Pi_0^{(N)} A\right), \quad (7.35)$$

where A is the following $N + 1 \times N + 1$ tridiagonal matrix

$$\begin{aligned}A &= \sum_{n=0}^N \binom{N}{n} |n\rangle\langle n| + \\ &- \frac{1}{2} \sum_{n=1}^N \sqrt{\binom{N}{n} \binom{N}{n-1}} (|n\rangle\langle n-1| + |n-1\rangle\langle n|). \quad (7.36)\end{aligned}$$

Within the symmetric subspace the group representa-

tion acts according to $U_\varphi^S = \sum_n e^{in\varphi} |n\rangle\langle n|$. The completeness condition (restricted to the symmetric subspace) implies that $(\Pi_0^{(N)})_n^n = 1$, while off-diagonal entries are arbitrary. Still, positivity condition requires that all off-diagonal elements have absolute values less or equal 1. Since off-diagonal terms in A will all contribute negatively to the final cost it is optimal to choose all off diagonal elements in the $\Pi_0^{(N)}$ operator to be 1. Hence the optimal

$$\Pi_0^{(N)} = \left(\sum_{n=0}^N |n\rangle \right) \left(\sum_{m=0}^N \langle m| \right) \quad (7.37)$$

and the resulting cost:

$$\begin{aligned} \bar{C}^{(N)} &= \frac{1}{2^{N-1}} \left(\sum_{n=0}^N \binom{N}{n} - \sum_{n=1}^N \sqrt{\binom{N}{n} \binom{N}{n-1}} \right) \\ &= 2 - \frac{1}{2^{N-1}} \sum_{n=1}^N \sqrt{\binom{N}{n} \binom{N}{n-1}}. \end{aligned} \quad (7.38)$$

One can check that in the limit of large N , $\bar{C}^{(N)} \rightarrow \frac{1}{N}$, see Problem 7.2, and hence we recover the identical asymptotic result as obtained within the frequentist approach.

Example 7.3 (Estimation of a completely unknown qubit state) We consider a model in which are given N copies of a completely unknown qubit state and our goal is to estimate it. Formally the state reads:

$$\rho_\psi^{(N)} = |\psi\rangle\langle\psi|^{\otimes N}, \quad |\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\varphi} \sin \frac{\theta}{2} |1\rangle, \quad (7.39)$$

where we parameterize the state of a qubit using Bloch sphere spherical angles. In order to think of this problem as a covariant estimation problem we may view $|\psi\rangle$ as obtained by rotating a fixed state $|0\rangle$ by a representation of the $SU(2)$ group. More precisely, since the initial state $|0\rangle$ will not change under rotations around the z axis, the relevant group will be $G = SU(2)/U(1)$.

$$|\psi\rangle = U_\psi |0\rangle = e^{i\varphi\sigma_z/2} e^{i\theta\sigma_y/2} |0\rangle. \quad (7.40)$$

The Haar measure in this case corresponds to the natural measure on the sphere so we will integrate over the states using

$$d\psi = \frac{1}{4\pi} d\theta d\varphi \sin \theta \quad (7.41)$$

As a cost function we choose:

$$C(\psi, \tilde{\psi}) = 4(1 - |\langle\psi|\tilde{\psi}\rangle|^2). \quad (7.42)$$

Note that this represents a distance derived from the fidelity measure, and for infinitesimally close states will reduce to standard squared distance on the sphere, see Problem... We have now formulated our problem as

a covariant estimation problem. Using (7.24) we have

$$\bar{C}^{(N)} = \int d\psi \operatorname{Tr} \left(\Pi_e |\psi\rangle\langle\psi|^{\otimes N} \right) 4(1 - |\langle\psi|0\rangle|^2), \quad (7.43)$$

which we need to minimize over Π_e , keeping in mind $\Pi_e \geq 0$, $\int d\psi U_\psi^{\otimes N} \Pi_e U_\psi^{\dagger \otimes N} = \mathbb{1}$. Let us rewrite the expression for the cost

$$\bar{C}^{(N)} = 4 \left[1 - \underbrace{\int d\psi \operatorname{Tr} \left(\Pi_e |\psi\rangle\langle\psi|^{\otimes N} \right) |\langle\psi|0\rangle|^2}_{\mathcal{F}} \right], \quad (7.44)$$

where we have introduced \mathcal{F} which may be viewed as the fidelity of estimation, which needs to be maximized. We can equivalently write \mathcal{F} as:

$$\mathcal{F} = \int d\psi \operatorname{Tr} \left(\Pi_e \otimes |0\rangle\langle 0| |\psi\rangle\langle\psi|^{\otimes N+1} \right), \quad (7.45)$$

where we have formally extended the space to $N+1$ qubit space in order to incorporate the cost function inside the trace operator. Note that

$$\mathcal{F} = \operatorname{Tr} \left(\Pi_e \otimes |0\rangle\langle 0| \underbrace{\int d\psi |\psi\rangle\langle\psi|^{\otimes N+1}}_A \right). \quad (7.46)$$

Let us study the properties of the A operator. This operator clearly is supported on the fully symmetric subspace of $N+1$ qubits and has trace 1. Moreover, this operator is invariant under the action of $U_\psi^{\otimes N+1}$, and since the fully-symmetric subspace carries the irreducible representation of the $U_\psi^{\otimes N+1}$ representation (with total angular momentum $j = (N+1)/2$), then by Schur lemma this operator must be proportional to identity on this subspace. The fully symmetric subspace of $N+1$ qubits has dimension $N+2$ and as a result:

$$A = \frac{1}{N+2} \mathbb{1}_{\mathcal{H}_S^{\otimes N+1}}. \quad (7.47)$$

We now need to find Π_e such that $\operatorname{Tr} \left(\Pi_e \otimes |0\rangle\langle 0| \mathbb{1}_{\mathcal{H}_S^{\otimes N+1}} \right)$ is maximal. We may restrict the Π_e operator to act solely on the symmetric subspace $\mathcal{H}_S^{\otimes N}$ as this the subspace where states $|\psi\rangle^{\otimes N}$ live. Let us denote $U_\psi^{j=N/2}$ to be the irreducible representation of $SU(2)$ acting on this subspace. Taking into account the completeness condition for Π_e :

$$\int d\psi U_\psi^{j=N/2} \Pi_e U_\psi^{j=N/2\dagger} = \mathbb{1}_{\mathcal{H}_S^{\otimes N}} \quad (7.48)$$

we see that $\operatorname{Tr} \Pi_e = N+1$. It is clear that in order to have the largest overlap between $\Pi_e \otimes |0\rangle\langle 0|$ and $\mathbb{1}_{\mathcal{H}_S^{\otimes N+1}}$, we would like to have $\Pi_e \otimes |0\rangle\langle 0|$ operator fully supported on $\mathcal{H}_S^{\otimes N+1}$. This will be so provided

we choose

$$\Pi_e = |0\rangle\langle 0|^{\otimes N}(N+1). \quad (7.49)$$

As a result we get $\mathcal{F} = \frac{N+1}{N+2}$ and finally

$$\bar{C}^{(N)} = 4 \left(1 - \frac{N+1}{N+2} \right) \quad (7.50)$$

Problem 7.1 Consider a qubit as a model of a two-level atom, where $|0\rangle$, $|1\rangle$ are respectively ground and excited state. Let us assume that we want to estimate transition frequency ω between the levels. We prepare the atom in state $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and evolve it subsequently for a known time t . As a result we obtain the state:

$$|\psi_\omega\rangle = e^{i\omega t\sigma_z/2}|\psi\rangle. \quad (7.51)$$

We assume that the prior distribution representing our prior knowledge about the frequency is gaussian:

$$p(\omega) = \frac{1}{\sqrt{2\pi\Delta^2\omega}} e^{-(\omega-\omega_0)^2/2\Delta^2\omega}, \quad (7.52)$$

where ω_0 and $\Delta^2\omega$ are the mean and the variance of the distribution respectively.

- Find the formula for the minimal Bayesian cost in this problem as a function of time of evolution t . Plot $\Delta^2\bar{\omega}/\Delta^2\omega$ as a function of t , which will show relative reduction of uncertainty as a result of the estimation procedure. Hint: In order to reduce you effort try to make use of the relation between the Bayesian cost and the Fisher information so you can use some of the results from the previous problem set.
- Determine the optimal evolution time for which the Bayesian cost is the lowest possible.
- For the optimal time provide the measurement and the values of estimated frequencies that result in the optimal estimation strategy.
- What would happen of somebody focused just on the quantum Fisher information of the state $|\psi_\omega\rangle$ in equation (7.51). At what conclusions he/she would arrive regarding the optimal evolution time using just the concept of Fisher information. Would these conclusions be sensible.

Problem 7.2 Prove that

$$\lim_{N \rightarrow \infty} N \cdot \frac{1}{2^{N-1}} \left(\sum_{n=0}^N \binom{N}{n} - \sum_{n=1}^N \sqrt{\binom{N}{n} \binom{N}{n-1}} \right) = 1. \quad (7.53)$$

Hint: Note that we can rewrite the left hand side as:

$$\lim_{N \rightarrow \infty} 2N \sum_n \frac{1}{2^N} \binom{N}{n} \left(1 - \sqrt{\frac{n}{N-n+1}} \right). \quad (7.54)$$

Now approximate the binomial distribution using a gaussian with mean $\langle n \rangle = N/2$ and $\Delta^2 n = N/4$. Introduce a new variable $x = n - N/2$ and expand the function to be averaged with the gaussian around $x = 0$. We know that x will be of order \sqrt{N} , so keep orders consistent while expanding $\left(1 - \sqrt{\frac{n}{N-n+1}} \right)$ up to orders $1/N$ (which means keeping also terms x^2/N^2).

Problem 7.3 In this chapter, it was proven that optimal Bayesowska strategy of estimating φ using N copies of state $|\psi_\varphi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + \exp(i\varphi)|1\rangle)$ and assuming the flat prior leads to the following formula for the minimal cost:

$$\bar{C}_{\text{opt}}^{(N)} = 2 - \frac{1}{2^{N-1}} \sum_{n=1}^N \sqrt{\binom{N}{n} \binom{N}{n-1}}, \quad (7.55)$$

where the cost function was chosen to be $C(\varphi, \tilde{\varphi}) = 4 \sin^2[(\varphi - \tilde{\varphi})/2]$.

Since the optimal strategy in general requires application of collective measurements on many particles, we would like to compare it with the performance of a simply single particle measurement and simple estimation based on the maximum likelihood estimator. Consider the following estimation strategy:

- We perform a measurement on a single particle described using 4 measurement operators: $\Pi_0 = \frac{1}{2}|+\rangle\langle+|$, $\Pi_1 = \frac{1}{2}|-\rangle\langle-|$, $\Pi_2 = \frac{1}{2}|+i\rangle\langle+i|$, $\Pi_3 = \frac{1}{2}|-i\rangle\langle-i|$, where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$, $|\pm i\rangle = (|0\rangle \pm i|1\rangle)/\sqrt{2}$. One can equivalently think about this measurement as a performing a projective measurement in basis $|+\rangle$, $|-\rangle$ with probability 1/2 and with probability 1/2 measuring in the $|+i\rangle$, $| - i\rangle$ basis. The measurement is performed subsequently on N particles. This way we obtain a sequence of measurement results $\vec{x} = (x_1, \dots, x_N)$, where $x_i \in \{0, 1, 2, 3\}$.

- b) Based on results \vec{x} we estimate phase $\tilde{\varphi}$ using the maximum likelihood estimator.
- c) We want to compare efficiency of this strategy with the optimal strategy. In order to do so, we fix some true value φ , and repeat the above steps approx 1000 times. For each realization we calculate the cost $C(\varphi, \tilde{\varphi})$. Since we want to compare the performance with the one optimal for the flat prior we repeat this procedure for different values of φ (e.g. 30 different values) uniformly placed within the interval $[0, 2\pi]$. We calculate the average cost $C^{(N)}$.
- d) We repeat the above strategy for different N and observe when we observe convergence to the value of the optimal Bayesian cost. We expect this to happen, since asymptotically the optimal cost behaves like $1/N$ and this we know we should be able to saturate using simple measurements. It will be instructive to identify N for which the advantage of the optimal strategy is the largest compared with the simple strategy, i.e. when $C^{(N)}/C_{\text{opt}}^{(N)}$ will be the largest

Part IV

Quantum metrology

Quantum metrology is the field of research where one aims at utilizing non-classical properties of light and atoms such as quantum coherence or entanglement in order to enhance the sensitivity of measuring devices. From a mathematical point of view quantum metrology represents next step in the quantum estimation theory as we will be effectively solving the problem of quantum channel estimation.

Consider a general quantum channel representing evolution of a quantum system, that depends on some unknown parameter θ :

$$\mathcal{E}_\rho(\rho) = \sum_i K_i^\theta \rho K_i^{\theta\dagger} = \rho_\theta, \quad (7.56)$$

where we have used the Kraus representation form, where K_i^θ are Kraus operators. Given such a model, the goal is not only to find the optimal measurement Π_x and estimator $\tilde{\theta}(x)$, that allows us to extract information on parameter θ from quantum state ρ_θ , but also to find the optimal input state ρ that maximizes the information that is available for extraction. In this sense we are adding one more element in the optimization task. Our goal will be to derive either strict limits or at least useful bounds on the achievable precision in such tasks. As before we will analyze the problem from two different perspectives: frequentist and Bayesian.

Chapter 8

Noiseless quantum metrology

We will start with the noiseless case first where the channel considered is unitary:

$$\rho_\theta = U_\theta \rho U^\dagger, \quad U_\theta = e^{i\theta H} \quad (8.1)$$

where H plays a role of the generator of the transformation U_θ and θ is the evolution parameter. In particular has we chose H to be the Hamiltonian of the system, the problem will be equivalent to the problem of estimating time of evolution of the quantum system.

It should be clear that without the loss of generality we may always restrict ourselves to pure state inputs. Any mixed state can be regarded as probabilistic mixtures of pure states, and it will never perform better than the best pure state that enters in the mixture. Moreover since we consider noiseless evolution the pure state will remain pure and hence we can restrict our considerations to pure states throughout this chapter:

$$|\psi_\theta\rangle = e^{iH\theta} |\psi\rangle. \quad (8.2)$$

8.1 Frequentist approach

Following the frequentist approach the natural way to identify the optimal protocols is to look for the input state $|\psi\rangle$ that yields the largest possible QFI of the output state $|\psi_\theta\rangle$. Since $|\psi_\theta\rangle = e^{iH\theta} |\psi\rangle$ then using (6.22) QFI reads:

$$F_Q = 4 (\langle \psi_\theta | H^2 | \psi_\theta \rangle - |\langle \psi_\theta | H | \psi_\theta \rangle|^2) = 4 (\langle \psi | H^2 | \psi \rangle - |\langle \psi | H | \psi \rangle|^2) = 4\Delta^2 H, \quad (8.3)$$

where we have used the fact that H commutes with U_θ . In order to find the bound on precision we therefore need to find the input state $|\psi\rangle$ for which the variance of the generator H is the largest possible.

Theorem 8.1 (The state maximizing QFI). Given the unitary parameter estimation problem, where $|\psi_\theta\rangle = e^{iH\theta} |\psi\rangle$, the input state that maximizes the QFI of the output state is:

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|\lambda_{\min}\rangle + |\lambda_{\max}\rangle), \quad (8.4)$$

where $|\lambda_{\min}\rangle, |\lambda_{\max}\rangle$ are eigenvectors of H corresponding to the minimal and maximal eigenvalues ($\lambda_{\min}, \lambda_{\max}$) respectively. The resulting maximal QFI reads:

$$F_Q = (\lambda_{\max} - \lambda_{\min})^2. \quad (8.5)$$

Proof. Let $H = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$ be the eigen-decomposition of H . First of all, notice that the variance is invariant with respect to the shift of all H eigenvalues by the same amount or in other words when we add to it something proportional to the identity: $\tilde{H} = H + \mu \mathbb{1} = \sum_i \tilde{\lambda}_i |\tilde{\lambda}_i\rangle\langle\tilde{\lambda}_i|$. Let us therefore assume that we have chosen μ so that for the new \tilde{H} $\tilde{\lambda}_{\min} = -\tilde{\lambda}_{\max}$ —the maximum and minimum eigenvalues are opposite to each other. This implies that $\forall_i |\tilde{\lambda}_i| \leq |\tilde{\lambda}_{\max}| = |\tilde{\lambda}_{\min}| = \tilde{\lambda}$. Let $|\psi\rangle = \sum_i c_i |\tilde{\lambda}_i\rangle$ be arbitrary state. According to (8.3) we can write the corresponding QFI:

$$F_Q = 4 \left(\sum_i |c_i|^2 \tilde{\lambda}_i^2 - \left(\sum_i |c_i|^2 \tilde{\lambda}_i \right)^2 \right). \quad (8.6)$$

Consider now the following chain of inequalities:

$$F_Q \leq 4 \sum_i |c_i|^2 \tilde{\lambda}_i^2 \leq 4 \sum_i |c_i|^2 \tilde{\lambda}^2 \leq 4\tilde{\lambda}^2. \quad (8.7)$$

The above inequalities are saturated if and only if we choose $|\psi\rangle = (|\tilde{\lambda}_{\max}\rangle + e^{i\xi}|\tilde{\lambda}_{\min}\rangle)/\sqrt{2} = (|\lambda_{\max}\rangle + e^{i\xi}|\lambda_{\min}\rangle)/\sqrt{2}$, where $e^{i\xi}$ is an arbitrary phase factor. Note also that $\tilde{\lambda} = (\tilde{\lambda}_{\max} - \tilde{\lambda}_{\min})/2 = (\lambda_{\max} - \lambda_{\min})/2$ which is a way to write the QFI in terms of eigenvalues of the original H . As a result we arrive at the optimal QFI as stated in the theorem. ■

8.1.1 N parallel channels

[FIGURE] We now move on to consider one of the most important questions in quantum metrology. Consider a situation where we deal with N channels that they can act in parallel on N probe systems via

$$U_{\theta}^{\otimes N} = e^{i\theta \sum_{k=1}^N H^{(k)}}, \quad (8.8)$$

where by $H^{(k)}$ denotes an operator that acts as H on the k -th probe system and as identity on all the others. Clearly, we can send into this setup N probes prepared in a product state $|\Psi\rangle = |\psi\rangle^{\otimes N}$. If additionally we choose $|\psi\rangle$ to be the state that maximizes the QFI in the single channel model, by additivity of the QFI we get the maximal QFI that can be achieved using product input states:

$$F_Q^{\text{product}} = N(\lambda_{\max} - \lambda_{\min})^2, \quad (8.9)$$

where as before λ_{\max} , λ_{\min} are maximal and minimal eigenvalues of H .

We now ask whether we can increase the QFI by considering entangled states at the input, and what is the maximal value of QFI in this case. To answer this question we see that we simply need to identify the optimal state realizing that the effective generator of the evolution is $H = \sum_k H^{(k)}$. The maximal and minimal eigenvalues are simply $\lambda_{\max}^{(N)} = N\lambda_{\max}$, $\lambda_{\min}^{(N)} = N\lambda_{\min}$, while the corresponding eigenvectors are $|\lambda_{\max}\rangle^{\otimes N}$, $|\lambda_{\min}\rangle^{\otimes N}$. As a result the optimal state is

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|\lambda_{\max}\rangle^{\otimes N} + e^{i\xi} |\lambda_{\min}\rangle^{\otimes N} \right), \quad (8.10)$$

where $e^{i\xi}$ is an arbitrary phase factor, whereas

the corresponding QFI reads:

$$F_Q^{\text{entangled}} = (N\lambda_{\max} - N\lambda_{\min})^2 = N^2(\lambda_{\max} - \lambda_{\min})^2, \quad (8.11)$$

and therefore is N time larger than in the product case. The resulting scaling of precision:

$$\Delta\varphi^{\text{product}} \geq \frac{1}{\sqrt{N}|\lambda_{\max} - \lambda_{\min}|}, \quad (8.12)$$

$$\Delta\varphi^{\text{entangled}} \geq \frac{1}{N|\lambda_{\max} - \lambda_{\min}|} \quad (8.13)$$

are often referred to as the standard (shot noise) and the Heisenberg scaling respectively.

In a special case, when we consider optical interferometry, we can think of N photons traveling through an interferometer as if through N independent qubit channels, where $H^{(k)} = \varphi\sigma_z^{(k)}/2$. In this case the optimally entangled state correspond to an equal superposition of N photons going one or another arm of the interferometer, which when written in photon number mode occupation notation reads:

$$|\psi\rangle = (|N, 0\rangle + |0, N\rangle)/\sqrt{2} \quad (8.14)$$

and is commonly referred to as the N00N state. The corresponding product and N00N state precision bounds read:

$$\Delta\varphi^{\text{product}} \geq \frac{1}{\sqrt{N}}, \quad \Delta\varphi^{\text{N00N}} \geq \frac{1}{N}. \quad (8.15)$$

8.1.2 General adaptive strategy

8.2 Bayesian approach

Identifying the optimal protocol within the Bayesian approach is much more challenging so let us restrict ourselves to phase estimation example. As usual, we assume flat prior distribution $p(\varphi) = 1/2\pi$, and the cost function $C(\varphi, \tilde{\varphi}) = 4\sin^2[(\varphi - \tilde{\varphi})/2]$. Consider a general N -photon two-mode input state

$$|\psi\rangle = \sum_n c_n |n\rangle |N - n\rangle = \sum_n c_n |n\rangle, \quad (8.16)$$

that after being sent through the phase imprinting operation U_φ reads:

$$|\psi_\varphi\rangle = U_\varphi|\psi\rangle = \sum_n c_n e^{in\varphi}|n\rangle. \quad (8.17)$$

Our goal is to minimize the Bayesian cost:

$$\bar{C} = \int \frac{d\varphi}{2\pi} \int \frac{d\tilde{\varphi}}{2\pi} \langle \psi_\varphi | \Pi_{\tilde{\varphi}} | \psi_\varphi \rangle C(\varphi, \tilde{\varphi}) \quad (8.18)$$

over the choice of input state $|\psi\rangle$ and measurement $\{\Pi_{\tilde{\varphi}}\}$. Since this is an example of a covariant estimation problem we know we can assume that $\Pi_{\tilde{\varphi}} = U_{\tilde{\varphi}} \Pi_0 U_{\tilde{\varphi}}^\dagger$ and the cost can be written effectively as:

$$\bar{C} = \frac{2}{\pi} \langle \psi | \int d\varphi U_\varphi^\dagger \Pi_0 U_\varphi \sin^2 \frac{\varphi}{2} | \psi \rangle. \quad (8.19)$$

The measurement completeness condition

$$\int \frac{d\tilde{\varphi}}{2\pi} U_{\tilde{\varphi}} \Pi_0 U_{\tilde{\varphi}}^\dagger = \mathbb{1}, \quad (8.20)$$

implies that again diagonal elements of Π_0 need to be 1. Similarly as in the phase estimation problem on product states $|\psi\rangle^{\otimes N}$, it can be shown that without loss of optimality (see Problem ???), the optimal Π_0 can be chosen to be a matrix with all entries equal 1:

$$\Pi_0 = |e\rangle\langle e|, \quad |e\rangle = \sum_{n=0}^N |n\rangle. \quad (8.21)$$

Consequently:

$$\begin{aligned} \bar{C} &= \frac{2}{\pi} \langle \psi | \int d\varphi \sum_{n,n'} e^{i(n-n')\varphi} |n\rangle\langle n'| \\ &\left(\frac{1}{2} - \frac{1}{4} e^{i\varphi} - \frac{1}{4} e^{-i\varphi} \right) | \psi \rangle = 2 \langle \psi | A | \psi \rangle, \end{aligned} \quad (8.22)$$

where A is a $(N+1) \times (N+1)$ tridiagonal matrix

$$A_{ij} = \begin{cases} 1 & i = j \\ -\frac{1}{2} & |i - j| = 1 \\ 0 & \text{otherwise} \end{cases}. \quad (8.23)$$

The eigendecomposition of this matrix can be determined analytically (see Problem ???) and its the minimum eigenvalue (which yields the minimal cost) and the corresponding eigenvector (the optimal state) read:

$$\bar{C} = 2 \left[1 - \cos \left(\frac{\pi}{N+2} \right) \right] \stackrel{N \rightarrow \infty}{\approx} \frac{\pi^2}{N^2} \quad (8.24)$$

$$|\psi\rangle = \sqrt{\frac{2}{N+2}} \sum_n \sin \left[\frac{(n+1)\pi}{N+2} \right]. \quad (8.25)$$

Interestingly, that we *do not* recover the frequentist solution in the asymptotic limit $N \rightarrow \infty$. The cost is by π^2 factor larger than in the frequentist approach and the character of the state is completely different than that of the N00N state. Note that the N00N state with its $2\pi/N$ periodicity is completely useless for estimation unless we know a priori we are within an parameter interval of the size $\sim \pi/N$. Had we used the N00N state in the Bayesian approach the effective cost would be identical to the one corresponding to random phase guessing! This is were the advantage of the Bayesian approach manifest itself. The π^2/N^2 is actually the *true* Heisenberg limit, i.e. the one that can be in principle reached if we are given N photons on which we want to run a quantum metrological scheme. The use of N00N states is not practical as for meaningful estimation we need to have prior knowledge on the order of π/N and hence be already in the Heisenberg limit regime!. Therefore, the conclusions of the frequentist approach are on valid under unrealistic requirements concerning prior knowledge about the parameter. Of course if we were allowed to repeat the estimation process many times this the frequentist approach would be more meaningful, but note that repeating experiment k times, we could expect to reach the variance $1/(N^2 k)$ using the total of kN photons, which will not correspond to the actual Heisenberg scaling $1/(kN)^2$ in terms of the total number of photons used.

8.2.1 General adaptive strategy

8.3 Phase vs. Frequency estimation

Problem 8.1 Consider the Mach-Zehnder interferometer, in which two single photons are simultaneously sent into the two input ports. After traveling through the first beam splitter the state becomes $|\psi\rangle = (|2, 0\rangle + |0, 2\rangle)/\sqrt{2}$ (Hong-Ou-Mandel interference) which is a superposition of two photons going the upper or the lower arm. The relative phase φ , related with interferometer arms difference, is then imprinted on the state, resulting in $|\psi_\varphi\rangle = (e^{2i\varphi}|2, 0\rangle + |0, 2\rangle)/\sqrt{2}$.

Try to find out what is the sensing potential of the $|\psi_\varphi\rangle$ state following two alternative approaches:

- a) Use the Quantum Cramer-Rao bound to provide the bound on maximal achievable precision of estimating φ .
- b) Apply the Bayesian approach, assuming flat prior distribution $1/(2\pi)$ and the cost function $C(\varphi, \tilde{\varphi}) = 4 \sin^2[(\varphi - \tilde{\varphi})/2]$ to find the minimal Bayesian cost.
- c) Compare the above two cases and draw conclusions

Problem 8.2 Kitaev

Chapter 9

Impact of decoherence

In previous chapters, while deriving fundamental precision limits in quantum metrology, we have restricted our attention to ideal decoherence-free models. In this chapter we want to understand the impact of decoherence on the achievable precision and in particular whether the Heisenberg scaling can be secured in presence of noise. Some of the physically relevant scenarios where methods will in particular find applications are optical interferometry in presence of losses and atomic interferometry in presence of dephasing.

Consider a quantum channel (CP map) Λ_θ , where θ is the parameter we want to estimate. Without loss of generality we may assume the input state to be pure (a mixture of states cannot perform better than the best state in the mixture) $|\psi\rangle$. However, the output state:

$$\rho_\theta = \Lambda_\theta(|\psi\rangle\langle\psi|) = \sum_i K_i^\theta |\psi\rangle\langle\psi| K_i^{\theta\dagger}, \quad (9.1)$$

will in general be mixed due to potentially non-unitary evolution encoded in Λ_θ — K_i^θ are Kraus operators, which all may in principle depend on θ . This makes the whole task of quantum metrology much more difficult as we cannot e.g. write simple formulas for the QFI of the output state not to mention optimize it analytically. For small dimensional systems, when trying to optimize input probe states, measurements and estimators we may of course resort to numerics. If, however, we want to derive efficiently computable bounds for parallel channel models involving many

particles or more generally adaptive schemes we have to invent ways to efficiently derive useful bounds even though exact optimization may not be possible.

In this chapter we will mainly follow the frequentist approach as it will be more effective in deriving the bounds, and typically will lead to results that are equivalent to the Bayesian ones in the limit of large number of probe systems. First, recall that QFI for a mixed state may be equivalently written as minimization over all purifications, see (6.52). [FIGURE - SCHEME] We may equivalently think of the purification of our quantum channel Λ_θ , so that

$$\rho_\theta = \text{Tr}_E \left(V_\theta |\psi\rangle\langle\psi| \otimes |0\rangle\langle 0| V_\theta^\dagger \right), \quad (9.2)$$

where V_θ is the unitary acting on extended $\mathcal{H}_S \otimes \mathcal{H}_E$ system + environment space. Therefore

$$|\Psi_\theta\rangle = V_\theta |\psi\rangle \otimes |0\rangle \quad (9.3)$$

is a purification of ρ_θ . Minimization over purification corresponds to minimization over equivalent unitary operations V_θ which may differ by an arbitrary unitary W_θ on E :

$$\tilde{V}_\theta = \mathbf{1} \otimes W_\theta \cdot V_\theta. \quad (9.4)$$

Kraus operators may be written in terms of V_θ in a standard way (see Chapter...) as $K_\theta^i = \langle i|V_\theta|0\rangle_E$ and hence different purifications may be viewed as inducing different equivalent

lent Kraus representations:

$$\tilde{K}_i^\theta = \sum_j (W_\theta)_{ij} K_j^\theta. \quad (9.5)$$

As will become clear further on, the key feature is that the unitary matrix W_θ is parameter dependent in general. We may now write the purification in terms of a given Kraus representation:

$$|\Psi_\theta\rangle = \sum_i K_i^\theta \otimes \mathbb{1} |\psi\rangle \otimes |i\rangle \quad (9.6)$$

and calculate the corresponding QFI:

$$F_Q(|\Psi_\theta\rangle) = 4 \left(\langle \dot{\Psi}_\theta | \dot{\Psi}_\theta \rangle - \left| \langle \dot{\Psi}_\theta | \Psi_\theta \rangle \right|^2 \right) = 4 \left(\langle \psi | \sum_i \dot{K}_i^{\theta\dagger} \dot{K}_i^\theta | \psi \rangle - \left| \sum_i \langle \psi | \dot{K}_i^{\theta\dagger} K_i^\theta | \psi \rangle \right|^2 \right) \leq 4 \langle \psi | \sum_i \dot{K}_i^{\theta\dagger} \dot{K}_i^\theta | \psi \rangle, \quad (9.7)$$

where $\dot{\cdot}$ denotes the derivative with respect to θ . Since the above inequality is valid for any purification (equivalent Kraus representation) we can write:

$$F_Q(\rho_\varphi) \leq 4 \min_{\{K_i^\theta\}} \langle \psi | \sum_i \dot{K}_i^{\theta\dagger} \dot{K}_i^\theta | \psi \rangle. \quad (9.8)$$

Finally, we may bound the QFI optimized over input states as:

$$\begin{aligned} \max_{|\psi\rangle} F_Q(\rho_\varphi) &\leq 4 \max_{|\psi\rangle} \min_{\{K_i^\theta\}} \langle \psi | \sum_i \dot{K}_i^{\theta\dagger} \dot{K}_i^\theta | \psi \rangle \leq \\ &\leq 4 \min_{\{K_i^\theta\}} \max_{|\psi\rangle} \langle \psi | \sum_i \dot{K}_i^{\theta\dagger} \dot{K}_i^\theta | \psi \rangle = \\ &= 4 \min_{\{K_i^\theta\}} \left\| \sum_i \dot{K}_i^{\theta\dagger} \dot{K}_i^\theta \right\|, \quad (9.9) \end{aligned}$$

where $\|\cdot\|$ is the operator norm, and in the above we have used the max-min inequality. As a result, we have bounded the achievable QFI on the output state using the objects defining the quantum channel itself. The utility of the above bound is not yet clear, unless

we are able to easily find Kraus representations that provide us with informative bounds.

9.1 Parallel channels

Let us now consider the parallel channel model, see Figure [FIGURE], where N particles prepared in a general, possibly entangled, state go through N independent channels Λ_θ , so that the total evolution is described by $\Lambda_\theta^{\otimes N}$. Physically, this represents the assumption that noise acting on each particle is uncorrelated and the parameter is encoded in the same way on each of the particles. When written in terms of Kraus operators the action of such a channel reads:

$$\Lambda_\theta^{\otimes N}(\cdot) = \sum_{i_1, \dots, i_N} K_{i_1}^\theta \otimes \dots \otimes K_{i_N}^\theta (\cdot) K_{i_1}^{\theta\dagger} \otimes \dots \otimes K_{i_N}^{\theta\dagger}. \quad (9.10)$$

Our goal is now to bound the QFI of the output state of N particles ρ_φ and try to capture the conditions under which the bound allows the QFI to scale like N^2 (Heisenberg scaling) or implies standard N scaling. The following theorem provides an efficiently computable bound that addresses this question.

Theorem 9.1. Given N parallel channels Λ_θ , where each channel is described using the set of Kraus operators K_i^θ , the QFI of the output state with respect to estimating the parameter θ for arbitrary input states of N particles is upper bounded as:

$$\max_{|\psi\rangle} F_Q(\rho_\theta) \leq 4 \min_{\{K_i^\theta\}} (N\|\alpha\| + N(N-1)\|\beta\|^2), \quad (9.11)$$

where the minimization is performed over all equivalent Kraus representations of Λ_θ , $\|\cdot\|$ denotes the operator norm and $\alpha = \sum_i \dot{K}_i^{\theta\dagger} \dot{K}_i^\theta$, $\beta = \sum_i \dot{K}_i^{\theta\dagger} K_i^\theta$.

Proof. Using the bound (9.9) we can write:

$$\begin{aligned} \max_{|\psi\rangle} F_Q(\rho_\theta) &\leq 4 \min_{\{K_i^\theta\}} \left\| \sum_{i_1, \dots, i_N} \left(K_{i_1}^\theta \otimes \dots \otimes K_{i_N}^\theta \right)^\dagger \right. \\ &\quad \left. \left(K_{i_1}^\theta \otimes \dots \otimes K_{i_N}^\theta \right) \right\|. \quad (9.12) \end{aligned}$$

Note that the minimization is performed over Kraus representations of a *single* channel. In principle we could assume more general Kraus representation where the channel independence property is not respected on the level of their purifications, but such a more general approach would not allow us to obtain efficiently computable bounds. Note, that restriction to subclass of purifications (Kraus representations) do not invalidate the bound but only involves the risk that the resulting bound may not be tight. Taking the derivative of the tensor product of Kraus operators we will get N terms where the derivative hits a given particle. We may then combine the resulting terms according to whether the derivative on the same particle in both parenthesis in the above expression or on different ones. Doing so we arrive at:

$$\max_{|\psi\rangle} F_Q(\rho_\theta) \leq 4 \left\| \sum_{i_1, \dots, i_N} \sum_{k=1}^N K_{i_1}^{\theta\dagger} K_{i_1}^\theta \otimes \dots \otimes \dot{K}_{i_k}^{\theta\dagger} \dot{K}_{i_k}^\theta \otimes \dots \otimes K_{i_N}^{\theta\dagger} K_{i_N}^\theta + \sum_{k \neq l} K_{i_1}^{\theta\dagger} K_{i_1}^\theta \otimes \dots \otimes \dot{K}_{i_k}^{\theta\dagger} K_{i_k}^\theta \otimes \dots \otimes K_{i_l}^{\theta\dagger} \dot{K}_{i_l}^\theta \otimes \dots \otimes K_{i_N}^{\theta\dagger} K_{i_N}^\theta \right\|.$$

Using the trace preservation condition $\sum_i K_i^{\theta\dagger} K_i^\theta = \mathbb{1}$ we may perform the sum whenever there is no derivative. Moreover, using the triangle inequality for the operator norm we arrive at:

$$\max_{|\psi\rangle} F_Q(\rho_\theta) \leq 4 \min_{\{K_i^\theta\}} \left(N \left\| \sum_i \dot{K}_i^{\theta\dagger} \dot{K}_i^\theta \right\| + N(N-1) \left\| \sum_i K_i^{\theta\dagger} K_i^\theta \right\|^2 \right), \quad (9.14)$$

which ends the proof. ■

The above theorem has profound consequences. Whenever we find a Kraus representation for which $\beta = 0$ this automatically implies that QFI cannot scale better than linearly with N and that the Heisenberg scaling is impossible.

Recall that different Kraus representations are defined via a unitary W_θ , see (??). Note also, that any parameter independent unitary will not affect the trace norms, as in this case dotted Kraus transform in the same way as undotted ones and the above expression is invariant under such a transformation. Moreover, since only first derivatives with respect to θ around some fixed point θ_0 are relevant

we may assume without loss of generality that our $W_\theta = Z_{\theta_0} e^{ih\theta}$, where h is some hermitian matrix, and Z_{θ_0} is some fixed unitary which does not influence the bound and may be set to $\mathbb{1}$. The formula for the bound may be therefore written in a more explicit form:

$$\max_{|\psi\rangle} F_Q(\rho_\theta) \leq 4 \min_h (N \|\alpha\| + N(N-1) \|\beta\|^2), \quad (9.15)$$

where

$$\alpha = \sum_i (\vec{K}^\theta + h\vec{K}^\theta)^\dagger (\vec{K}^\theta + h\vec{K}^\theta) \quad (9.16)$$

$$\beta = \sum_i (\vec{K}^\theta + h\vec{K}^\theta)^\dagger \vec{K}^\theta \quad (9.17)$$

where \vec{K}^θ is the vector of Kraus operators $(\vec{K}^\theta)_i = K_i^\theta$.

Example 9.1 (Atomic interferometry in the presence of dephasings) Consider a phase estimation model using N two-level atoms that experience decoherence in the form of uncorrelated dephasing. In other words, we consider an arbitrary N -qubit state that is acted upon by $\Lambda_\varphi^{\otimes N}$, where $\Lambda_\varphi(\cdot) = \Lambda(U_\varphi \cdot U_\varphi^\dagger)$, $U_\varphi = \exp(i\varphi/2)|0\rangle\langle 0| + \exp(-i\varphi/2)|1\rangle\langle 1|$ is unitary phase imprinting operations while Λ is the pure dephasing process that can be written using two Kraus operators: $K_0 = \mathbb{1} \sqrt{(1+\eta)/2}$, $K_1 = \sigma_z \sqrt{(1-\eta)/2}$ —the single qubit Bloch sphere is being shrunk in the xy plane by a factor of η .

Let us consider the following equivalent Kraus representation:

$$\tilde{K}_0^\varphi = \cos(\chi\varphi) K_0 U_\varphi - i \sin(\chi\varphi) K_1 U_\varphi, \quad (9.18)$$

$$\tilde{K}_1^\varphi = \cos(\chi\varphi) K_1 U_\varphi - i \sin(\chi\varphi) K_0 U_\varphi \quad (9.19)$$

parameterized by χ . For this Kraus representation α and β operators read respectively:

$$\alpha = \frac{\mathbb{1}}{4} \left(1 + \chi^2 + \chi \sqrt{1 - \eta^2} \right) \quad (9.20)$$

$$\beta = \sigma_z \left(\frac{1}{2} + \chi \sqrt{1 - \eta^2} \right). \quad (9.21)$$

We see that we can to set $\beta = 0$ by choosing $\chi = -\frac{1}{2\sqrt{1-\eta^2}}$ and as a result $\alpha = \mathbb{1} \frac{\eta^2}{4(1-\eta^2)}$. This implies the bound:

$$F_Q \leq \frac{N\eta^2}{1-\eta^2}, \quad \Delta\varphi \geq \frac{\eta}{\sqrt{(1-\eta^2)N}}. \quad (9.22)$$

This can be compared with the performance when using optimal product states $|\psi\rangle = [(|0\rangle + |1\rangle)/\sqrt{2}]^{\otimes N}$

which gives $F_Q = N\eta^2$ (see Problem). While the Heisenberg scaling is not possible to be protected in presence of decoherence there is a chance for a constant factor improvement of estimation precision given by $\sqrt{1-\eta^2}$. As will be demonstrated in Chapter ?? this enhancement is indeed possible to obtain.

9.2 General adaptive protocols

9.3 Constrained protocols

Problem 9.1 Optimal product state in presence of dephasing.

Problem 9.2 Interferometer with losses. $\Delta\varphi \geq \sqrt{\frac{1-\eta}{\eta N}}$.

Chapter 10

Practical protocols

Chapter 11

General Markovian noise models

Chapter 12

Quantum error-correction

Chapter 13

Waveform estimation

Chapter 14

Adiabatic quantum metrology

Part V
And beyond

Chapter 15

Quantum discrimination

Chapter 16

Quantum communication

Chapter 17

Quantum computing

Chapter 18

Quantum thermodynamics