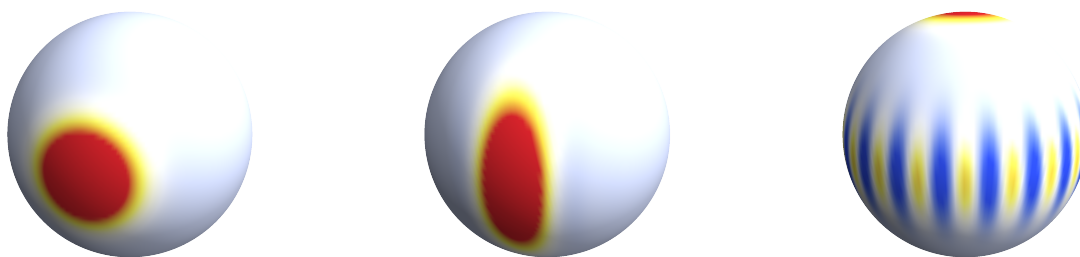


Doctoral Thesis

Heisenberg Limit beyond Quantum Fisher Information

Wojciech Górecki



Supervisor

dr hab. Rafał Demkowicz-Dobrzański

Referees

prof. dr hab. Dariusz Chruściński, Nicolaus Copernicus University in Toruń

prof. dr hab. Michał Horodecki, University of Gdańsk

dr hab. Emilia Witkowska, Polish Academy of Sciences

Thesis submitted in

the Chair of Quantum Optics and Atomic Physics Institute of Theoretical Physics,
Faculty of Physics, University of Warsaw

in fulfilment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

Heisenberg Limit beyond Quantum Fisher Information

Wojciech Górecki

The Heisenberg limit provides a fundamental bound on the achievable estimation precision with a limited number of N resources used (e.g., atoms, photons, etc.). Using entangled quantum states makes it possible to scale the precision with N better than when resources would be used independently. Consequently, the optimal use of all resources involves accumulating them in a single execution of the experiment. Unfortunately, that implies that the most common theoretical tool used to analyze metrological protocols - quantum Fisher information (QFI) – does not allow for a reliable description of this problem, as it becomes operationally meaningful only with multiple repetitions of the experiment.

In this thesis, using the formalism of Bayesian estimation and the minimax estimator, I derive asymptotically saturable bounds on the precision of the estimation for the case of noiseless unitary evolution. For the case where the number of resources N is strictly constrained, I show that the final measurement uncertainty is π times larger than would be implied by a naive use of QFI. I also analyze the case where a constraint is imposed only on the average amount of resources, the exact value of which may fluctuate (in which case QFI does not provide any universal bound for precision). In both cases, I study the asymptotic saturability and the rate of convergence of these bounds.

In the following part, I analyze the problem of the Heisenberg limit when multiple parameters are measured simultaneously on the same physical system. In particular, I investigate the existence of a gain from measuring all parameters simultaneously compared to distributing the same amount of resources to measure them independently. Using two examples – the measurement of multiple phase shifts in a multi-arm interferometer and the measurement of three magnetic field components – I show the existence of qualitative differences between the results obtained using Bayesian estimation/minimax estimator and those resulting from the use of QFI. I also derive a lower bound on the achievable precision of the measurement in the general case (not always saturable).

ACKNOWLEDGEMENT

First and foremost, I would like to sincerely thank my promoter Rafal Demkowicz-Dobrzanski for his unlimited support throughout my doctoral studies, combined at the same time with leaving great freedom for development and strong motivation to achieve results.

My sincere thanks to Dominic Berry and Howard Wiseman, who, long before me, began to address the Heisenberg limit beyond quantum Fisher information, whose publications inspired me and in collaboration with whom I wrote the first of the articles that form the basis of this thesis. I would also like to thank Sisi Zhou and Liang Jiang, with whom I also had the pleasure of working in the early years of my Ph.D. on an independent project. Contact with these four at such an outstanding level so early in my career made a huge impression on me and influenced the quality of my future work.

I would also like to thank Stanisław Kurdziałek, Francesco Albarelli, and Lorenzo Maccone, with whom I worked directly in Warsaw and Pavia, for all the substantive discussions and a great atmosphere of cooperation, which fueled each other.

I thank Jacek Krajczok, Tomasz Cheda, Albert Rico, and Paweł Kurzyński for all the conversations at the intersection of quantum information, algebra, and statistics, which allowed me to become convinced that the field of my research and the knowledge gained during my Ph.D. is not as hermetic as it might seem.

I also want to thank everyone who supported me early in my scientific career and helped me find the direction I want to go in theoretical physics. In particular, I would like to thank my high school teacher, Elżbieta Zawistowska, as well as the supervisors of my internships during my early studies, Kazimierz Rządewski and Krzysztof Pawłowski.

I would also like to thank everyone who has supported me in my personal life over the years.

During my doctorate, I was financially supported by the National Science Center (Poland) grant No. 2016/22/E/ST2/00559 and the National Science Center (Poland) Grant No. 2020/37/B/ST2/02134. I was also supported by the Foundation for Polish Science (FNP) via the START scholarship.

LIST OF PUBLICATIONS

Publications this thesis is based on:

W. Górecki, R. Demkowicz-Dobrzański, *Multiparameter quantum metrology in the Heisenberg limit regime: Many-repetition scenario versus full optimization*, [Physical Review A 106, 012424 \(2022\)](#)

W. Górecki, R. Demkowicz-Dobrzański, *Multiple-Phase Quantum Interferometry: Real and Apparent Gains of Measuring All the Phases Simultaneously*, [Physical Review Letters 128, 040504 \(2022\)](#)

W. Górecki, R. Demkowicz-Dobrzański, H.M. Wiseman, D.W. Berry, *π -Corrected Heisenberg Limit*, [Physical Review Letters 124, 030501 \(2020\)](#)

Publications related to quantum metrology cited in this thesis:

S. Kurdziałek, **W. Górecki**, F. Albarelli, R. Demkowicz-Dobrzański, *Using adaptiveness and causal superpositions against noise in quantum metrology*, [arXiv:2212.08106 \(2022\)](#)

W. Górecki, A. Riccardi, L. Maccone, *Quantum metrology of noisy spreading channels*, [Physical Review Letters 129, 240503 \(2022\)](#)

R. Demkowicz-Dobrzański, **W. Górecki**, M. Guță, *Multi-parameter estimation beyond quantum Fisher information*, [Journal of Physics A: Mathematical and Theoretical 53, 363001 \(2020\)](#)

W. Górecki, S. Zhou, L. Jiang, R. Demkowicz-Dobrzański, *Optimal probes and error-correction schemes in multi-parameter quantum metrology*, [Quantum 4, 288 \(2020\)](#)

Other publications:

J. Kopyciński, M. Łebek, **W. Górecki**, K. Pawłowski, *Ultrawide dark solitons and droplet-soliton coexistence in a dipolar Bose gas with strong contact interactions*, [arXiv:2207.03817 \(2022\)](#), accepted in Physical Review Letters

J. Kopyciński, M. Łebek, M. Marciniak, R. Ołdziejewski, **W. Górecki**, K. Pawłowski, *Beyond Gross-Pitaevskii equation for 1D gas: quasiparticles and solitons*, [SciPost Physics 12, 023 \(2022\)](#)

W. Golletz, **W. Górecki**, R. Ołdziejewski, K. Pawłowski, *Dark solitons revealed in Lieb-Liniger eigenstates*, [Physical Review Research 2, 033368 \(2020\)](#)

- R. Ołdziejewski, **W. Górecki**, K. Pawłowski, K. Rzażewski, *Strongly correlated quantum droplets in quasi-1D dipolar Bose gas*, [Physical Review Letters 124, 090401 \(2020\)](#)
- R. Ołdziejewski, **W. Górecki**, K. Pawłowski, K. Rzażewski, *Roton in a few-body dipolar system*, [New Journal of Physics 20, 123006 \(2018\)](#)
- R. Ołdziejewski, **W. Górecki**, K. Pawłowski, K. Rzażewski, *Many-body solitonlike states of the bosonic ideal gas*, [Physical Review A 97, 063617 \(2018\)](#)
- W. Górecki**, K. Rzażewski, *Electric dipoles vs. magnetic dipoles-for two molecules in a harmonic trap*, [EPL \(Europhysics Letters\) 118, 66002 \(2017\)](#)
- W. Górecki**, K. Rzażewski, *Making two dysprosium atoms rotate—Einstein-de Haas effect revisited*, [EPL \(Europhysics Letters\) 116, 26004 \(2016\)](#)
- R. Ołdziejewski, **W. Górecki**, K. Rzażewski, *Two dipolar atoms in a harmonic trap*, [EPL \(Europhysics Letters\) 114, 46003 \(2016\)](#)

ABBREVIATIONS AND NOTATION

MSE	mean square error,
RMSE	root mean square error,
FI	Fisher information,
QFI	quantum Fisher information
CR	Cramér-Rao inequality,
N	the total number of resources used to estimate the parameter.
n, k	n is number of resources used in single trial, k is number of trials.
HS	Heisenberg scaling ($\text{MSE} \sim 1/N^2$),
POVM	positive operator-valued measure,
SLD	symmetric logarithmic derivative,
ML estimator	maximum likelihood estimator,
θ	parameter to be measured,
$\tilde{\theta}(\cdot)$	estimator,
H	generator of the evolution of the whole system,
Λ	generator of the evolution of a single particle,
$\mathcal{E}(\cdot)$	single particle quantum channel,
$\mathcal{H}_S$	single particle Hilbert space,
$\mathcal{H}_A$	ancillary system Hilbert space,

All formulas cited from other papers are translated to this notation, with additional footnote comments, where needed.

TABLE OF CONTENTS

ABSTRACT	2
ACKNOWLEDGEMENT	3
LIST OF PUBLICATIONS	4
ABBREVIATIONS AND NOTATION	6
CHAPTER 1 : INTRODUCTION	9
1.1 Heisenberg limit in quantum metrology	9
1.2 Major problems discussed in the thesis	9
1.3 Outline of the thesis	10
CHAPTER 2 : QUANTUM METROLOGY AND THE HEISENBERG LIMIT	12
2.1 Heisenberg limit by an example	12
2.2 General formulation of the channel estimation problem	15
2.3 The role of observable and optimality of projective measurements	17
2.4 Average energy case and the inability to overcome the Heisenberg scaling	18
2.5 Entangled-parallel and sequential-adaptive metrological schemes	20
CHAPTER 3 : FISHER INFORMATION AND CRAMÉR-RAO BOUND	24
3.1 Classical Cramér-Rao bound	24
3.2 Example – two arms interferometer	25
3.3 Asymptotic normality of maximum likelihood estimator	26
3.4 Quantum Cramér-Rao bound	28
3.5 Quantum metrology – unitary evolution	30
3.6 Infinite QFI for average energy constraints	32
3.7 General theorem for a noisy estimation	33
Example	37
CHAPTER 4 : OBTAINABLE HEISENBERG LIMIT IN SINGLE PARAMETER UNITARY ESTI- MATION	39
4.1 Bayesian and minimax costs	39
4.2 Optimality of covariant measurements	41
4.3 Interferometer with a fixed total number of photons	42
4.4 Interferometer with a fixed average number of photons	45

4.5	π -corrected Heisenberg limit	46
	Asymptotic saturability	51
	Intuitive interpretation	52
4.6	Average energy constraint	53
CHAPTER 5 : MULTIPARAMETER METROLOGY		55
5.1	Foundations	55
5.2	Fisher information matrix and multiparameter Cramér-Rao bound	55
	Adaptive scheme maximizing QFI	58
	General theorem for noisy multiparameter estimation within QFI formalism	62
5.3	Bayesian and minimax approaches	62
	Covariant problem	63
CHAPTER 6 : HEISENBERG LIMIT IN MULTIPARAMETER METROLOGY		68
6.1	Introduction	68
6.2	Asymptotic equivalence of the global and local minimax cost	69
6.3	Multiple-phase interferometer	72
6.4	Magnetic field with unknown direction – $SU(2)$ estimation	76
6.5	General bound for precision in multiparameter unitary evolution	80
6.6	Magnetic field in two spatially separated points	82
6.7	Summary	84
CHAPTER 7 : CONCLUSION AND OUTLOOK		85
APPENDICES		86
BIBLIOGRAPHY		92

LIST OF ILLUSTRATIONS

FIGURE 2.1	Wigner function for the estimation of the spin's rotation	12
FIGURE 2.2	Interferometer	19
FIGURE 2.3	Metrological schemes	22
FIGURE 3.1	Maximum likelihood estimator	26
FIGURE 3.2	Saturating QFI with average energy constrain	34
FIGURE 6.1	Multiple-phase interferometer metrology scheme	73

CHAPTER 1

INTRODUCTION

1.1. Heisenberg limit in quantum metrology

In classical mechanics, all parameters characterizing a physical system can, in principle, be measured to any accuracy, but, in practice, the imperfections of the measurement apparatus introduce certain fluctuations in the results. Repeated measurement reduces the uncertainty of the estimation of a parameter, such that the mean square error of the estimator (MSE) scales inversely to the number of these repetitions (the so-called shot noise limit).

In contrast to the above, the formalism of quantum mechanics rigorously defines achievable precision, as the act of measurement itself is not deterministic but probabilistic. As a result, the question of the fundamental limit on the achievable precision of the measurement of a given parameter becomes a well-posed problem without the need to specify a particular measurement procedure. Moreover, quantum mechanics opens up the prospect of achieving a better scaling of measurement accuracy with the number of resources used in the experiment (which can be understood as the number of atoms, number of photons, total time, total energy, etc.). Using appropriate entangled quantum states [Horodecki et al. [2009], Mitchell et al. [2004]] (or sequential multiple passes [Higgins et al. [2007]]) makes it possible to decrease MSE quadratically with the number of resources. Such scaling has been named after Werner Heisenberg by Holland and Burnett [Holland and Burnett [1993]], where they referred to number-phase uncertainty relation in [Heitler [1954]].

The issue of the feasibility of reaching the Heisenberg limit is intensively discussed in the current literature [Giovannetti et al. [2006], Paris [2009], Giovannetti et al. [2011], Toth and Apellaniz [2014], Demkowicz-Dobrzanski et al. [2015], Schnabel [2017], Degen et al. [2017], Pezzè et al. [2018], Pirandola et al. [2018], Niezgoda et al. [2019]]. At the same time, in some cases, there are ambiguities related to the interpretation of results obtained by different methods.

1.2. Major problems discussed in the thesis

The existence of Heisenberg scaling implies that when the only constraint is the total number of resources, the optimal strategy is to accumulate them all in a single execution of the experiment. However, the results obtained with most of the theoretical tools commonly used in quantum metrology, such as Fisher's quantum information (QFI) and the related Cramér-Rao (CR) inequality, have an unambiguous interpretation only in the limit of many repetitions. As a result, the actual limit on achievable precision is not very well explored, and there are inconsistent (sometimes even contradictory) statements on this topic in the current literature.

The most widely discussed example in the quantum metrology literature is the issue of phase shift estimation in an interferometer. Back in the 1990s, it was noted that when the value of the shift is

initially completely unknown (it can take any value in the interval $[0, 2\pi[$), the minimum uncertainty, obtainable using N photons, is π times larger than would result from naive use of QFI [Luis and Peřina [1996], Buřek et al. [1999], Berry and Wiseman [2000]]. In 2011, an argument was put forward [Hayashi [2011]] justifying that, in the limit of N going to infinity, this difference is independent of the size of the $[0, 2\pi[$ interval. In this thesis, I generalize this observation to the case of any unitary evolution generator with a bounded spectrum. Furthermore, I show that the limit on precision is also valid for more advanced adaptive metrology protocols. Finally, I do not restrict myself to the limit of N going to infinity but also discuss the case of finite N . These results have already been published in [Górecki et al. [2020]].

A second major issue related to the misunderstandings arising from the misuse of quantum Fisher information in the context of the Heisenberg limit is the precision achievable when only the expectation value of the total energy is bounded. In such a situation, the QFI can be unconstrainedly large. Some authors have postulated the existence of protocols that allow for better scaling than Heisenberg scaling [Rivas and Luis [2012], Anisimov et al. [2010]]; however, these works were criticized rather quickly [Pezzé [2013]]. In 2012, it was shown for the general case that the MSE cannot decrease faster than quadratically with the mean energy [Giovannetti et al. [2012]], but the derived bound was not saturable. In the thesis, I analyze this problem by establishing the saturable bound and investigating the rate of convergence. These results are based on the methodology and observations from my papers [Górecki et al. [2020], Górecki and Demkowicz-Dobrzański [2022a]], but they were not published before in explicit form.

Finally, I examine the issue of the Heisenberg limit in the context of multi-parameter metrology. One of the aspects addressed is the comparison of the efficiency of the combined measurement of all parameters simultaneously with the situation when they are measured separately. The analysis of such an issue necessitates a well-defined procedure for resource allocation between the parameters, which has been treated radically differently in various publications [Humphreys et al. [2013a], Yousefjani et al. [2017]] leading to contradictory results regarding the existence of a gain from performing a joint measurement. This paper discusses the issue of resource allocation and makes a comprehensive comparison of the results obtained using the two paradigms (QFI and Bayesian estimation theory/minimax approach). The presented results have been published in [Górecki and Demkowicz-Dobrzański [2022a,b]].

1.3. Outline of the thesis

In Chapter One, I present, by way of example, the fundamental problems that arise when attempting to analyze the Heisenberg limit using concepts restricted to local estimation, such as QFI or the error propagation formula. I also discuss the state of the art of the precision bound for average energy constraint. I formulate the problem of a quantum channel estimation, and I introduce a distinction between entangled-parallel and adaptive-sequential metrological schemes.

In Chapter Two, I discuss the Fisher quantum information formalism, highlighting its strengths (such as the extensively developed theory concerning metrology in the presence of noise) while emphasizing its limitations in the context of Heisenberg limit analysis.

In Chapter Three, I analyze the operationally achievable Heisenberg limit under constraints on the total amount of resources available in the system or on the average energy. In doing so, I use the formalism of Bayesian estimation and minimax estimator. In this chapter, I present the results from [Górecki et al. [2020], Górecki and Demkowicz-Dobrzański [2022a]].

In Chapter Four, I go further to the multiparameter case, introducing basic concepts in both QFI formalism and Bayesian/minimax one. I reproduce the proof of optimality of entangled-parallel strategy for the covariant problems [Chiribella et al. [2008b]], which I will need in the next chapter.

Finally, in Chapter Five, I discuss the Heisenberg limit in multiparameter case, presenting the results from [Górecki and Demkowicz-Dobrzański [2022a,b]].

CHAPTER 2

QUANTUM METROLOGY AND THE HEISENBERG LIMIT

"Everything should be made as simple as possible, but not simpler," Albert Einstein was once supposed to have said. Keeping in mind this rule, before introducing broader mathematical formalism, I would like to briefly explain the essence of the main problem I will discuss in this thesis by introducing a simple but expressive example.

2.1. Heisenberg limit by an example

Consider the problem of estimating the strength of a weak magnetic field oriented in vertical direction z , sensed by atoms with internal magnetic moment – spin $1/2$. Since action of the field on the atoms results in rotating them around the z axis via unitary operation $U_\theta = e^{-i\theta J_z}$ (where J_z is the spin z component¹), the problem is equivalent to estimating the angle of the rotation θ , see Fig. 2.1(a). The aim is to perform the whole procedure so that the mean squared error of the estimator $\Delta^2 \tilde{\theta}$ is minimized.

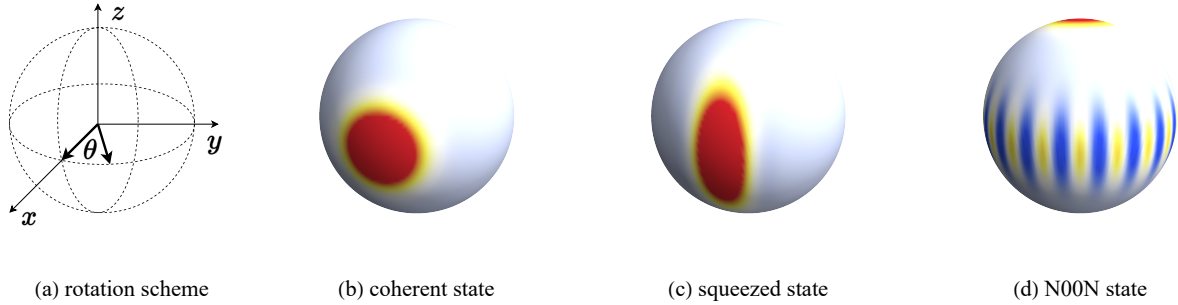


Figure 2.1: (a) The vector of internal magnetic moment (spin) is rotated by an angle θ around vertical axis z . (b-d) Due to fundamental laws of quantum mechanics, all three components of the spin vector cannot be simultaneously well-defined. Wigner function [Jing et al. [2019], Davis et al. [2021]] expressing these uncertainties are plotted for three different quantum states of $N = 20$ atoms

The simplest and most intuitive strategy is to orient the atom's spin in direction x and, after rotation, measure the y component of the spin J_y . As for small rotation the value of $\langle J_y \rangle \approx \frac{1}{2}\theta$, one can simply estimate the value of parameter as $\tilde{\theta} = 2\langle J_y \rangle$. From the error propagation formula:

$$\Delta^2 \tilde{\theta} = \frac{\Delta^2 J_y}{\left| \frac{d\langle J_y \rangle}{d\theta} \right|^2}. \quad (2.1)$$

¹For simplicity of further formulas, dividing by the constant factor $\hbar$ is incorporated into the definition of J_z , in such way, that J_z is dimensionless.

As for a single particle, angular momentum operator J_y is simply Pauli matrix multiplied by factor one-half $\frac{1}{2}\sigma_y$, the variance of such measurement's result is finite and equal $\Delta^2 J_y = \frac{1}{4}$, which leads to $\Delta^2 \tilde{\theta} = 1$ (in radians). If one repeats measurement on N independent atoms and then estimates the angle value from the averaged result, the variance of the measurement results will decrease by a factor $\frac{1}{N}$. At the same time, the denominator will remain unchanged, leading at the end to $\Delta^2 \tilde{\theta} = \frac{1}{N}$. Such scaling with the amount of resources used is called standard scaling or shot noise limit.

The question is whether one may obtain better scaling by collectively using all N atoms. The first guess would be to coherently orient all atoms to the same direction x , such that it may be effectively seen as a single particle of spin $N/2$, see Fig. 2.1(b). Then even a small rotation by an angle θ will lead to N times bigger changes of the value of the collective spin $\langle J_y \rangle \approx \frac{1}{2}N\theta$. Unfortunately, it would also increase the fluctuation of the y component itself $\Delta^2 J_y = N$. As a result, these two effects will cancel each other, leading again to the shot noise limit, i.e., the same precision could be obtained by measuring each atom separately and averaging the result.

May this problem be overcome in quantum mechanics by choosing another state for sensing? The laws of quantum mechanics do not allow us to cancel the uncertainties completely. The famous Heisenberg uncertainty principle, which, in its most canonical form, establishes the relationship between momentum and position variances $\Delta x \Delta p \geq \frac{\hbar}{2}$ (where $\Delta x := \sqrt{\Delta^2 x}$ is a root mean squared error (RMSE)), in general bounds the product of variances of any two observables by the expectation value of their commutator [Schrödinger [1930], Robertson [1934]]:

$$\Delta J_y \Delta J_z \geq \frac{1}{2} |\langle [J_y, J_z] \rangle|, \quad (2.2)$$

which, in our case, is simply equal to the value of the spin component in x direction $|\langle J_x \rangle|$. Therefore, even keeping J_x still large, we have some freedom to decrease the variance $\Delta^2 J_y$, by squeezing the state along y axes. It simultaneously increases the uncertainty in z direction $\Delta^2 J_z$, but this effect does not affect our estimation procedure, see Fig. 2.1(b).

Is there any fundamental bound for the precision of estimating angle θ using N spins $1/2$? To answer this question, note that as the evolution of the system is simply rotating around z axis $e^{-i\theta J_z}$, the speed of change of expectation value $\langle J_y \rangle$ is also determined by the expectation value of the commutator:

$$\frac{d\langle J_y \rangle}{d\theta} = i\langle [J_y, J_z] \rangle. \quad (2.3)$$

Substituting $\frac{1}{i} \frac{d\langle J_y \rangle}{d\theta}$ for $\langle [J_y, J_z] \rangle$ in Eq. (2.2) and dividing both sides by ΔJ_z and $\frac{d\langle J_y \rangle}{d\theta}$ one obtain

following bound [Mandelstam and Tamm [1991]]:

$$\Delta^2 \tilde{\theta} \geq \frac{1}{4\Delta^2 J_z}. \quad (2.4)$$

Quite surprisingly, even though in our procedure we decided to obtain information about the rotation angle by measuring the spin component in the y -direction, neither $\langle J_y \rangle$ nor $\Delta^2 J_y$ appears in the final formula. Indeed, the whole reasoning remains valid for any other measurement that extracts information about θ .

It strongly suggests that looking for the state optimal for measuring rotation, we should choose the one which maximizes the variance $\Delta^2 J_z$. It would be the famous Greenberger–Horne–Zeilinger (GHZ) state (in the context of quantum optics also called $|N00N\rangle$ state) – a quantum superposition of situation, where all N atoms' spins are oriented up and 0 are oriented down, with the situation, where 0 is oriented up, and N is oriented down – $|N00N\rangle \propto \frac{1}{\sqrt{2}}(|\uparrow\rangle^{\otimes N} + e^{iN\theta}|\downarrow\rangle^{\otimes N})$, for which $\Delta^2 J_z = N^2$. We see a significant improvement compared to the initially proposed strategy – the bound Eq. (2.4) indicate that precision could scale quadratically with N , which we will call Heisenberg scaling (HS). However, looking at the graphical representation of this state Fig. 2.1(d), one may find an interesting property – in addition to the two strongly outlined non-pole areas (corresponding directly to the downward and upward spins independently), one can also see dense fringes at the equator, resulting from the quantum interference of the two states. The fringes are very narrow so that even when rotated by a small angle, the arrangement of the fringes changes significantly, making it possible to observe even a minimal change precisely. On the other hand, the cyclic nature of the distribution of the fringes means that when rotated through an angle $\theta = \frac{2\pi}{N}$ (or multiple thereof), their distribution becomes identical again (which corresponds to the phase factor $e^{iN\theta}$ in the formula above). As the results in such a scenario, we can estimate the value of the parameter only up to the term $+\frac{m2\pi}{N}$ ($m \in \mathbb{Z}$). It means that to use $|N00N\rangle$ state (potentially leading to $\Delta\tilde{\theta} = \frac{1}{N}$), from the very beginning, we need to know the exact value of the parameter with accuracy $2\pi/N$. This fact makes the single usage of $|N00N\rangle$ extremely poor strategy, as it does not significantly increase initial knowledge.

Another way to see how the $|N00N\rangle$ cannot be effective for estimating general unknown angle θ is information theory. Note that even if $|N00N\rangle$ consist N atoms, independently on the exact value of θ it remain into two-dimensional Hilbert space spanned by $\{\frac{1}{\sqrt{2}}(|\uparrow\rangle^{\otimes N} + |\downarrow\rangle^{\otimes N}), \frac{1}{\sqrt{2}}(|\uparrow\rangle^{\otimes N} - |\downarrow\rangle^{\otimes N})\}$. Therefore, from the point of view of the information capacity, it should be treated as a simple qubit and hence can transmit at most a single bit of information. That shows that even if it can record information on a tiny angle change in a readable way, it is impossible to record information on an angle change over a wide range.

When the true parameter value is guaranteed to lay in a small neighborhood of size $\sim \pi/N$ of some

fixed value θ_0 , the $|N00N\rangle$ is extremely useful for detecting arbitrary small changes of θ . If one is then able to repeat procedure k times to estimate the angle from the averaged result, they will obtain $\Delta^2\tilde{\theta} \approx \frac{1}{kN^2}$ (which would lead to a significant update of the initial knowledge).

However, such a strategy became suboptimal from the view of total resources $k \cdot N$ used, as precision scales only linearly with k . Consequently, the inequality Eq. (2.4) constitutes the valid bound for the estimation precision, but it is not necessarily saturable. Therefore, it does not answer the question about fundamentally optimal precision of estimation in a situation where only the total amount of resources is restricted. This problem cannot be solved solely using an error propagation formula or any other tool based on local formalism, i.e., the one which only considers the first derivatives of the functions around the working point (as Fisher information formalism). This thesis aims to analyze it going beyond this methodology.

The problem may be formulated more abstractly in the general form – for any unitary evolution $U_\theta = e^{i\theta H}$ governed by the hermitian generator of transformation H , where some peculiar state is used for sensing this parameter, the variance of the estimator will be bounded by the inverse of the variance of the H calculated on this state:

$$\Delta^2\tilde{\theta} \geq \frac{1}{4\Delta^2 H}. \quad (2.5)$$

2.2. General formulation of the channel estimation problem

Given unitary quantum channel acting on the system, depending on the unknown parameter θ : $\rho_\theta = U_\theta \rho U_\theta^\dagger$. The aim is to get an information about the parameter θ by preparing a proper input state ρ , acting the channel and performing a general measurement – positive operator-valued measure (POVM) $\{M_x\}$ ($\forall_x M_x \geq 0$, $\sum_x M_x = \mathbb{1}$) on the output state ρ_θ . The probability of obtaining the result x when the parameter value is equal θ is given by the Born rule $p(x|\theta) = \text{Tr}(M_x \rho_\theta)$. The measurement may be repeated many times, resulting in the sequence of outcomes $\mathbf{x} = \{x_1, \dots, x_k\}$. At last, one needs to choose the estimator $\tilde{\theta}(\mathbf{x})$ assigning an estimated parameter value to a sequence of results. The measure of the efficiency of the estimator is the MSE:

$$\Delta^2\tilde{\theta} = \sum_{\mathbf{x}} p(\mathbf{x}|\theta) (\tilde{\theta}(\mathbf{x}) - \theta)^2, \quad (2.6)$$

where $p(\mathbf{x}|\theta) = \prod_{i=1}^k p(x_i|\theta)$. In the case when the mean value of the estimator indicates the correct value of the parameter (the estimator is unbiased):

$$\sum_{\mathbf{x}} p(\mathbf{x}|\theta) \tilde{\theta}(\mathbf{x}) = \theta, \quad (2.7)$$

the MSE is equal to the variance of the estimator. In such cases, I will use these words interchangeably if it does not lead to ambiguity.

One should notice that for a given measurement and estimator, the value of MSE Eq. (2.6) depends on the exact value of θ , so the problem of finding optimal metrology strategy is not well defined until we define a range of parameter value for which the estimator is to work correctly. In principle, we would like the estimator to return actual value independently of its value. However, this condition in many situations may be too strict or even impossible to satisfy for a finite number of repetitions. Moreover, we often have some approximated knowledge about the parameter value, and we are interested in local estimation around a specific point $\theta \approx \theta_0$, so it is not needed for the estimator to work well very far from this value. Still, if one demands from estimator to work correctly only in single point θ_0 , then trivial constant estimator $\forall_{\mathbf{x}} \tilde{\theta}(\mathbf{x}) = \theta_0$ would lead to zero cost, while it extracts no information from the measurement. To avoid such pathological constructions, heuristically, we would say we require that the estimator works correctly, at least in a certain immediate neighborhood of θ_0 . As further analysis will show, different approaches to the mathematical formalization of this condition will lead to different results.

One way is to impose the condition that estimator is **locally unbiased** around point θ_0 , i.e., Eq. (2.7) as well as its first derivative are satisfied at point θ_0 :

$$\begin{aligned} \sum_{\mathbf{x}} p(\mathbf{x}|\theta_0) \tilde{\theta}(\mathbf{x}) &= \theta_0, \\ \sum_{\mathbf{x}} \left. \frac{dp(\mathbf{x}|\theta)}{d\theta} \right|_{\theta=\theta_0} \tilde{\theta}(\mathbf{x}) &= 1. \end{aligned} \tag{2.8}$$

Note that the above condition allows for an exact rederivation of the error propagation formula Eq. (2.1) for the situation, where the parameter θ is estimated with the use of some observable $A = \sum_a a|a\rangle\langle a|$. Indeed, for measurement $\{M_a\} = \{|a\rangle\langle a|\}$ direct minimization of the estimator variance with locally unbiased condition results in $\tilde{\theta}(a) = \theta_0 + a \left(\left. \frac{d\langle A \rangle}{d\theta} \right|_{\theta=\theta_0} \right)^{-1}$ and $\Delta^2 \tilde{\theta} = \Delta^2 A / \left(\left. \frac{d\langle A \rangle}{d\theta} \right|_{\theta=\theta_0} \right)^2$.

As shown in the discussed example, the analysis performed only with local unbiasedness conditions may not be meaningful when the experiment is performed only once. However, the concept of local unbiasedness is a very efficient tool when the experiment is repeated many times. The relation between the minimal cost obtainable with local unbiasedness condition for single measurement realization vs. the cost with global unbiasedness condition in the limit of many repetitions will be discussed in Ch. 3. Still, as argued at the beginning, we are interested in the optimal usage of whole resources, which demands using all of them in a single realization so we will need another theoretical tool.

The problem of full minimization of the MSE of the estimator, with restriction only on the total

amount of resources N will be discussed in [Ch. 4](#) with the usage of two alternative approaches. In the Bayesian approach, one introduces a priori distribution of the value of parameter $p(\theta)$ and as the figure of merit, consider the [Eq. \(2.6\)](#) averaged over this probability. Alternatively, in the minimax approach, the finite size set of possible values of parameter Θ is under consideration, and the figure of merit is [Eq. \(2.6\)](#) maximized over $\theta \in \Theta$.

To make the results from [Ch. 3](#) and [Ch. 4](#) somehow comparable and to avoid confusion, I introduce a proper notation for the number of used resources. By N I will always understand the total amount of resources used. In the scenario with many experiment's repetitions k , I will denote the number of resources used in a single trial as n , such that $N = k \cdot n$. In this context, one may think about the results from [Ch. 3](#) as the ones obtained with optimization over protocols using in total N resources, with additional constraints for the maximal amount of resources used in a single trial n (so the measurement is repeated $k = N/n$ times).

2.3. The role of observable and optimality of projective measurements

The attentive reader may note that in [Sec. 2.1](#) we discussed the direct estimation of a parameter from the mean value of the observables, whereas in [Sec. 2.2](#) we consider independent optimization over both the measurement and the estimator itself. Therefore, one may ask whether the bound [Eq. \(2.5\)](#) is still valid with such an extension.

In the following, I will show that, in fact, for the case where the measurement is not repeated many times (which corresponds to optimal use of resources), optimization over the observable itself is equivalent to optimization over the general measurement and the estimator.

In the literature, this fact has been derived independently in different formalisms (see [[Macieszczak et al. \[2014\]](#)] for the Bayesian approach). Below I want to show that for all introduced approaches (locally unbiased estimators, Bayesian or minimax approach), it may be seen as a direct consequence of a single matrix inequality.

For a given measurement $\{M_x\}$ and estimator $\tilde{\theta}(x)$, the mean value of the estimator for any true value of the parameter θ may be written as the expectation value of the observable A defined as follow:

$$\langle \tilde{\theta} \rangle = \sum_x \text{Tr}(\rho_\theta M_x) \tilde{\theta}(x) = \text{Tr} \left(\rho_\theta \underbrace{\sum_x M_x \tilde{\theta}(x)}_A \right). \quad (2.9)$$

Now I will argue that such a procedure may only be improved if one measures the observable directly (i.e., performs the projection onto its eigenvectors $\{|a\rangle\langle a|\}$) and then estimates the value of θ from the average value of A (formally: attributes to each of $|a\rangle\langle a|$ the estimator $\tilde{\theta}'(a)$, such that $A = \sum_a \tilde{\theta}'(a) |a\rangle\langle a|$).

To show that, I construct the positive operator of the following form (its positivity comes from $\forall_x M_x \geq 0$):

$$0 \leq \sum_x \left((\tilde{\theta}(x) - \theta) \mathbb{1} - (X - \mathbb{1}\theta) \right) M_x \left((\tilde{\theta}(x) - \theta) \mathbb{1} - (X - \mathbb{1}\theta) \right) = \sum_x M_x (\tilde{\theta}(x) - \theta)^2 - (A - \mathbb{1}\theta)^2 \quad (2.10)$$

(note that no assumptions about unbiasedness have been used here; the last equality comes directly from $\sum_x M_x = \mathbb{1}$ and $A = \sum_x M_x \tilde{\theta}(x)$). From that, we have:

$$\forall_\theta \quad \Delta^2 \tilde{\theta} = \sum_x \text{Tr}(\rho_\theta M_x) (\tilde{\theta}(x) - \theta)^2 \geq \sum_x \text{Tr}(\rho_\theta (A - \mathbb{1}\theta)^2), \quad (2.11)$$

where the last inequality is tight if all M_x are the projections onto eigenstates of A .

As [Eq. \(2.11\)](#) holds for any strategy $\{M_x\}$, $\tilde{\theta}(x)$ at any θ , it remains valid if one restricts to locally unbiased estimators, as well as if one averages both sides with arbitrary $p(\theta)$ or take the maximum over $\theta \in \Theta$.

Still, it should be stressed that above we have considered the single parameter estimation, with single measurement realization, where the figure of merit is MSE. In more general cases, separating estimator from measurement may still be useful and necessary, i.e., for the situations where:

- one considers many repetitions of the experiment, so the measurement is local, but the estimator may depend on all measurement results jointly.
- the cost function other than MSE is under consideration.
- more than one parameter is to be estimated.

All these cases will also appear in further chapters.

2.4. Average energy case and the inability to overcome the Heisenberg scaling

[Eq. \(2.5\)](#) establishes a fundamental lower bound (not necessarily saturable) for the estimator's variance concerning the variance of the evolution generator for the sensing state. As shown in the discussed example, if the total number of involved resources is N , the generator variance may scale like N^2 , leading to quadratic improvement to the classical strategy. The natural question is, what happens if one imposes constraints only on the average number of resources, not the maximum one?

The problem was extensively discussed in 2010s in the context of optical interferometry. Consider two-arms interferometer, with unknown phaseshift θ in the upper arm, see [Fig. 2.2](#). Let us label the modes connected to both arms by $\hat{a}, \hat{a}^\dagger$ and $\hat{b}, \hat{b}^\dagger$; the corresponding generator of the evolution is then $H = \hat{a}^\dagger \hat{a}$. Note that for the case when the number of photons N is fixed the problem is mathematically exactly equivalent to the one discussed in [2.1](#), if in the latter one we consider only

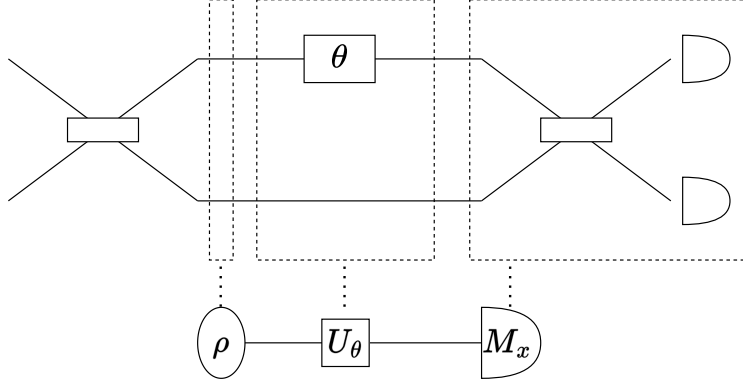


Figure 2.2: The simplest scheme of the optical interferometer. The incoming state of light is divided between the sensing (upper) and the reference (lower) arm. After going through the interferometer, the light again passes through a beam splitter. In the end, photon-counting detectors are applied to each output arm, where the statistics of this photon counting depend on the phase shift θ in the sensing arm. Generally, the procedure may be optimized over preparing the input state ρ and applying more advantage measurement.

the fully symmetric subspace of states. Indeed, if we appropriately identify the up $|\uparrow\rangle$ and down $|\downarrow\rangle$ spin states with lower and upper arms in the interferometer, the single-particle evolution generator is the same (up to, physically irrelevant, global phase).

However, an interesting effect may be observed if the constraint is imposed only on the average number of photons $\langle H \rangle = \bar{N}$. Then one may easily construct the state with arbitrary large $\Delta^2 H$ – for example, if the probability of finding exactly k photons in the upper arm decreases like $\propto \frac{1}{k^3}$, the average converges to constant, while the variance converges to infinity. In some papers, this fact was interpreted as the opportunity to overcome this Heisenberg scaling $\frac{1}{N^2}$, and a few concrete potential protocols have been proposed. Further analysis showed that they lead to strongly biased results or require a large number of repetitions, such that at the end, the scaling with a whole amount of used resources is not better than $\frac{1}{N^2}$ (see also [Sec. 3.6](#) for the discussion by the example within Fisher information formalism).

The universal solution for such a class of problems has been proposed in [[Giovannetti et al. \[2012\]](#)], where the authors presented the bound based on quantum speed limit² (e.i. how fast the fidelity between states $|\psi\rangle$ and $e^{i\theta H}|\psi\rangle$ may decrease with increasing θ) [[Margolus and Levitin \[1998\]](#), [Giovannetti et al. \[2003\]](#)]. They have shown that for any reasonable estimator

$$\Delta^2 \tilde{\theta} \geq \max \left[\frac{\kappa^2}{(\langle H \rangle - E_g)^2}, \frac{\gamma^2}{\Delta^2 H} \right], \quad (2.12)$$

²Similar reasoning has been proposed earlier in [[Zwierz et al. \[2010\]](#)], however, the derivation of the bound was invalid [[Zwierz et al. \[2011\]](#)].

where E_g is the ground state energy (the minimal eigenvalue of H), while κ and γ are constant (not given analytically). By "reasonable," they only assume that the region where the estimator works well should be at least two times bigger than resulting $\Delta\tilde{\theta}$.

At a similar time, the bound for the particular case of this problem (the estimation of completely random phase shift in interferometer) has been derived [Hall et al. [2012], Hall and Wiseman [2012]], based on entropic uncertainty relations [Białynicki-Birula and Mycielski [1975], Hall et al. [2012]]. This bound is tighter but less general, while qualitative consequences are the same.

The above works close the discussion on the optimal scaling with the average amount of resources and the impossibility of beating the Heisenberg scaling. However, due to the not-saturability of mentioned bounds, they do not provide an answer to the question regarding the multiplicative constant in the minimal obtainable cost.

2.5. Entangled-parallel and sequential-adaptive metrological schemes

So far, we have discussed the problem of unitary evolution governed by generator H , which allows for formulating the bound in terms of its average value or its variance. Especially for the system with non-interactive particles or linear optics, $\langle H \rangle$ typically scales linearly with a number of particles (or photons), leading to Heisenberg scaling of the precision of the form $\sim \frac{1}{N^2}$. Let us focus on such a case to introduce a notation for the rest of the thesis.

For the sake of simplicity and homogeneity of notation, let me now ignore the indistinguishability of particles (bosons or fermions) and treat them as distinguishable (the problem will be discussed in more detail in specific cases, if necessary). It is obvious that for distinguishable particles, the estimation would be easier (because a larger states' space would be available), so all the bounds derived for such a case are still valid for bosons and fermions (while saturation may depends on which case is considered).

Let me name the Hilbert space of the single particle in the discussed system by $\mathcal{H}_S$ and the corresponding single particle evolution generator by $\Lambda \in \text{Lin}(\mathcal{H}_S)$ (where $\text{Lin}(\mathcal{H})$ denotes the space of linear operators acting on $\mathcal{H}$). The N -particle evolution is therefore governed by:

$$H = \sum_{i=1}^N \mathbb{1}^{\otimes(i-1)} \otimes \Lambda \otimes \mathbb{1}^{\otimes(N-i)}, \quad e^{i\theta H} = \left(e^{i\theta \Lambda} \right)^{\otimes N}. \quad (2.13)$$

In most of the analyses presented in this thesis, I will not significantly go beyond the problem of unitary evolution. However, to provide a more realistic context, at some point, I will also discuss the recent results about the noisy case (obtained typically within QFI formalism). To have a unified framework, therefore, I denote a single particle quantum gate as $\mathcal{E}_\theta(\rho) = \rho_\theta$, which for unitary evolution is given as $\mathcal{E}_\theta(\rho) = e^{i\theta \Lambda} \rho e^{-i\theta \Lambda}$ (the more general case of arbitrary completely

positive trace preserving (CPTP) map [Chruściński and Kossakowski [2009]] will be discussed later (Sec. 3.7). Let us now consider a different way of using N quantum gates.

In the simplest scenario, the experimentalist uses N uncorrelated particles, lets them evolve, and makes the measurement on each of them (see Fig. 2.3(a)):

$$\rho_\theta = \mathcal{E}_\theta(\rho), \quad p(\{x_1, \dots, x_N\}|\theta) = \prod_{i=1}^N \text{Tr}(M_{x_i} \rho_\theta). \quad (2.14)$$

In such an approach, the MSE will scale as $\sim 1/N$, due to the standard statistical properties. Alternatively, one may use general entangled N -particle state $\rho \in \text{Lin}(\mathcal{H}_S^{\otimes N})$, for which the evolution is given by acting the product of single-particle channels $\mathcal{E}_\theta^{\otimes N}(\rho)$. In general, such state may be additionally entangled with some external ancillary system $\mathcal{H}_A$, and then for $\rho \in \text{Lin}(\mathcal{H}_S^{\otimes N} \otimes \mathcal{H}_A)$ we have (see Fig. 2.3(b)):

$$\rho_\theta = (\mathcal{E}_\theta^{\otimes N} \otimes \mathbb{1})(\rho) \quad (2.15)$$

which in principle may allow for $\sim 1/N^2$ scaling of MSE.

Keeping this formalism, we may consider an even more general scenario, where one can use the channel $\mathcal{E}_\theta(\cdot)$ N times in a completely arbitrary way, including multi-times acting on a single particle (or group of entangled particles), with arbitrary unitary control and/or partial measurement between. The most general adaptive scheme³ may be written as (see also Fig. 2.3(b)):

$$\rho_\theta = V_N \circ (\mathcal{E}_\theta \otimes \mathbb{1}) \circ \dots \circ V_1 \circ (\mathcal{E}_\theta \otimes \mathbb{1}) \circ \rho, \quad (2.16)$$

where ρ is a single-particle state entangled with the ancilla $\rho \in \text{Lin}(\mathcal{H}_S \otimes \mathcal{H}_A)$, $V_i \circ \rho$ is a shortcut for $V_i \rho V_i^\dagger$, while V_i are unitary controls acting jointly on the system and ancilla. Note that this equation covers all mentioned cases. Indeed, any actions of the channel on multi-particle states may be simulated with the above equation by applying proper SWAP operators in unitary control V_i . Furthermore, the partial measurement may also be simulated, assuming that the ancillary system is sufficiently large.

Formulating the number of resources in terms of quantum gates used in the protocol is helpful, as it allows for generalizing the problem for arbitrary channel estimation and easily includes more sophisticated sequential adaptive strategies. Moreover, by taking proper limits, it may also be applied to the problems of continuous time evolution [Górecki et al. [2020]].

However, it should be noted that it does not cover all potential metrological issues. For example, for

³In modern literature, even more general schemes appear, where indefinite causal order of acting of gates is considered [Zhao et al. [2020], Liu et al. [2022]]. However, its physical interpretation is not fully clear, and therefore this case will not be discussed in this thesis.

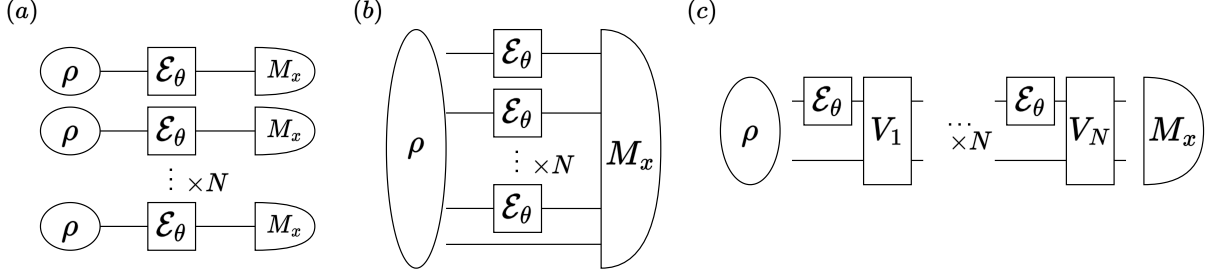


Figure 2.3: Three metrological schemes of usage N quantum gates to estimate the parameter θ . a) gates are used independently, which may lead only to standard scaling of the MSE $\sim 1/N$. b) entangled-parallel scheme: all N gates act jointly in a parallel way on some entangled state (extended by ancilla if needed). c) sequential-adaptive scheme: most general scheme, including sequential acting of the gate N times with unitary control between and entanglement with arbitrary large ancilla.

the systems with two-body interaction, the evolution of the state of N particles cannot be written as Eq. (2.15). Moreover, in such case, the energy may scale as N^2 with the number of particles, so the bound Eq. (2.12) allows for quadratic scaling $\Delta^2 \tilde{\theta} \sim \frac{1}{N^4}$ [Beltrán and Luis [2005], Boixo et al. [2007]]. Another example may be amplitude estimation of trapped ions [Wolf et al. [2019]] or light [Górecki et al. [2022]], where in the latter case, the channel may increase the number of photons. Finally, it does not provide an accurate description of the problem where the noise acting of different probes is mutually correlated [Chabuda et al. [2020]]. Such cases will not be this thesis's topic; we should still be aware of their existence.

Another point should be highlighted here. During reformulating the problem from scaling with the energy to scaling with the number of gates used, we have restricted ourselves to the problems of the well-defined number of gates (representing photons etc.). On the other hand, previously cited bound Eq. (2.12) allows for non-well-defined energy (unbounded from above).

In [Demkowicz-Dobrzański et al., 2015, Section 5.3.2.] it was pointed out that for the problems in linear optics, it is much more reasonable to formulate the bounds in terms of the maximal amount (not average) of the resources. The reason is that, in practice, we cannot measure the superposition of the states with a different number of photons, so they will be no more useful than corresponding mixtures. Then we can use the argument that for most reasonable models, the minimal variance obtainable with a fixed amount of resources N is a convex function of N (typically $\sim \frac{1}{N}$ or $\sim \frac{1}{N^2}$), so the optimal strategy with the mean amount of resources will be the one with the well-defined amount of resources.

Still, the problem of optimal strategy with constraining only on the average amount of resources may be relevant for other models (moreover, it will be helpful in analyzing a specific problem of multiparameter estimation in Sec. 6.3). Recently, some approaches to formulating the problem of

sequential adaptive schemes with energy restriction appear in literature [Zhao et al. [2020]], but the general formulation is not yet established. This case, however, will not be explored in this thesis. While talking about the bound for the estimator's variance with the constraint on mean energy, I will restrict myself only to the entangled-parallel scheme, which will be discussed in [Sec. 4.4](#), [Sec. 4.6](#).

CHAPTER 3

FISHER INFORMATION AND CRAMÉR-RAO BOUND

In this chapter, I introduce the formalism based on Fisher information and Cramér-Rao inequality, stating the fundamental bounds on estimation precision when the experiment is repeated many times. My emphasis is on pointing out how the results obtained by this method should be interpreted and what their operational significance is. At the end of the chapter, I also present current developments in analyzing the estimation problem in the presence of noise.

3.1. Classical Cramér-Rao bound

Given probability distribution $p(x|\theta)$ of the measurement result x depended on the value of unknown parameter $\theta \in \Theta$. The aim is to estimate the value of the parameter from the measurement outcomes. Then for any estimator, locally unbiased at $\theta = \theta_0$, i.e., the one satisfying:

$$\langle \tilde{\theta} \rangle \Big|_{\theta=\theta_0} = \sum_x p(x|\theta_0) \tilde{\theta}(x) = \theta_0 \quad (3.1)$$

$$\frac{d\langle \tilde{\theta} \rangle}{d\theta} \Big|_{\theta=\theta_0} = \sum_x \frac{dp(x|\theta)}{d\theta} \Big|_{\theta=\theta_0} \tilde{\theta}(x) = 1, \quad (3.2)$$

its variance is bounded from below by Cramér-Rao bound (CR):

$$\Delta^2 \tilde{\theta} \geq \frac{1}{F(p)}, \quad F(p) = \sum_x \frac{1}{p(x|\theta)} \left(\frac{dp(x|\theta)}{d\theta} \right)^2, \quad (3.3)$$

where $F(p)$ is classical Fisher information (FI) of the probability distribution $p(x|\theta)$.

CR bound is typically derived using Cauchy Schwarz inequality; it may also be obtained by direct minimization over estimators with locally unbiased constraint using the Lagrange multipliers method. In the latter method, one would obtain:

$$\tilde{\theta}^{\text{l.u.}}(x) = \frac{1}{F(p)} \frac{1}{p(x|\theta)} \frac{dp(x|\theta)}{d\theta}. \quad (3.4)$$

Note that while the above estimator is by construction always locally unbiased at point θ_0 , we have no knowledge about the size of the neighborhood around θ_0 , where it works well, which in general makes FI and CR bound not very useful for the case when the experiment is performed only once.

Directly from the definition, FI is convex, i.e., for any two probability distributions $p_1(x|\theta)$, $p_2(x|\theta)$

$$F(c_1 p_1 + c_2 p_2) \leq c_1 F(p_1) + c_2 F(p_2), \quad (3.5)$$

and additive, i.e., for the joint probability of two variables of the product form $p_{12}(x_1, x_2|\theta) = p_1(x_1|\theta)p_2(x_2|\theta)$ we have

$$F(p_{12}) = F(p_1) + F(p_2). \quad (3.6)$$

Especially for the k -length sequence of measurement results $\mathbf{x} = \{x_1, \dots, x_k\}$, wherein each repetition's outcome is drawn independently, so $p(\mathbf{x}) = p(\{x_1, \dots, x_k\}) = p(x_1)p(x_2)\dots p(x_k)$, the FI is simply k times bigger than FI for single realization, so:

$$\Delta^2 \tilde{\theta} \geq \frac{1}{k \cdot F(p)}. \quad (3.7)$$

It is consistent with a simple observation that one may always choose as the collective estimator the average of single-result estimators $\tilde{\theta}(\mathbf{x}) = \frac{1}{k} \sum_i \tilde{\theta}(x_i)$, without violating locally unbiasedness condition. However, such a procedure would be a poor choice in the limit of large repetitions number k , as it would not solve the problem of biasedness far from the point θ_0 .

It turns out that in the limit of large k , [Eq. \(3.7\)](#) may be saturated in all points $\theta \in \Theta$ by the same estimator. That is the maximum likelihood estimator (ML estimator), i.e., the one associating to any sequence of the measurement outcomes the value of θ , for which the probability of obtaining this sequence is the biggest possible:

$$\tilde{\theta}_{\text{ML}}(\mathbf{x}) = \arg \max_{\theta} p(\mathbf{x}|\theta). \quad (3.8)$$

For such estimator, [[Vaart, 1998](#), 5.3. Asymptotic Normality] [[Nagaoka and Hayashi, 2005](#), CHAPTER 7]

$$\begin{aligned} \forall \theta \quad \lim_{k \rightarrow \infty} \langle \tilde{\theta}_{\text{ML}}(\mathbf{x}) \rangle &= \theta \\ \lim_{k \rightarrow \infty} k \Delta^2 \tilde{\theta}_{\text{ML}}(\mathbf{x}) &= \frac{1}{F[p(\cdot|\theta)]}, \end{aligned} \quad (3.9)$$

where in principle the value of $F[p(\cdot|\theta)]$ may depend on the value of parameter.

As the precise formulation of converging ML estimator to the normal distribution is crucial for understanding ambiguities regarding the interpretation of FI in the context of HL, for completeness, in [Sec. 3.3](#), I attach the proof of optimality of ML estimator. Before, let me discuss it with a simple but expressive example.

3.2. Example – two arms interferometer

Let us go back to the problem of estimating the phase shift in two arms interferometer [Fig. 2.2](#) and assume that in each trial, only one photon is put into the system. The probabilities of clicking of

each detector are given as follows:

$$p(0|\theta) = (1 - \sin(\theta))/2, \quad p(1|\theta) = (1 + \sin(\theta))/2. \quad (3.10)$$

The problem of estimating θ is therefore equivalent to the analyzing the problem the unfair coin of probability $q = (1 + \sin(\theta))/2$. To find the maximum of $p(\{x_1, \dots, x_k\}|\theta) = q^{\sum x_i} (1 - q)^{k - \sum x_i}$ over θ , it is easier to deal with its logarithm:

$$\log(p(\{x_1, \dots, x_k\}|\theta)) = \left(\sum x_i\right) \log(q) + \left(k - \sum x_i\right) \log(1 - q). \quad (3.11)$$

Then we have:

$$\frac{dp(\{x_1, \dots, x_k\}|\theta)}{dq} = \frac{\sum x_i}{q} - \frac{k - \sum x_i}{1 - q} = \frac{\sum x_i - kq}{q(1 - q)} = 0 \Rightarrow \frac{\sum x_i}{k} = q, \quad (3.12)$$

which leads to the exact formula for the ML estimator:

$$\tilde{\theta}_{\text{ML}}(\{x_1, \dots, x_k\}) = \arcsin\left(1 - 2\frac{\sum x_i}{k}\right). \quad (3.13)$$

The expected value of this estimator's indication and MSE (depending on the number of performed trials and the actual value of the parameter θ) are presented in Fig. 3.1. For comparison, the performances of locally unbiased estimators Eq. (3.4) at point $\theta_0 = 0$ and $\theta_0 = 1$ (for $k = 100$) are also presented.

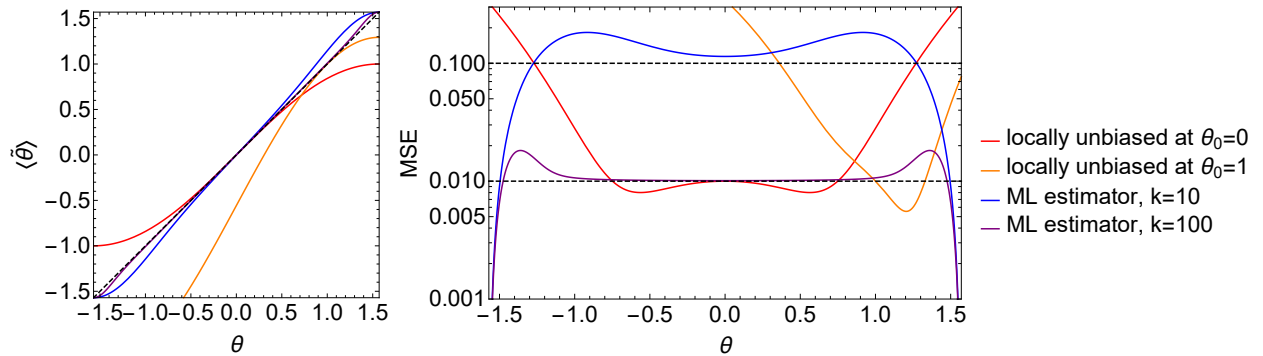


Figure 3.1: Performance of ML estimator and locally unbiased estimator for the problem of estimation of the phase shift in the interferometer. (Left panel) the expectation value of the estimator. (Right panel) MSE dependence on the actual value of θ . One may see that a locally unbiased estimator works well only in the closest neighborhood of θ_0 , while ML already for 100 trials returns reasonable values for all θ and it is close to converging CR bound (dashed horizontal lines).

3.3. Asymptotic normality of maximum likelihood estimator

To formulate theorem precisely, let me remind two types of convergence of random variables.

The sequence of random variables Y_k converge in distribution to random variable Y iff for any y cumulative distribution function of Y_k converge to cumulative distribution functions of Y :

$$Y_k \xrightarrow{d} Y \Leftrightarrow \forall_y \lim_{k \rightarrow \infty} p(Y_k \leq y) = p(Y \leq y). \quad (3.14)$$

The sequence of random variables Y_k converge in probability to a constant c iff:

$$Y_k \xrightarrow{p} c \Leftrightarrow \forall_\epsilon \lim_{k \rightarrow \infty} p(|Y_k - c| \geq \epsilon) = 0. \quad (3.15)$$

Theorem. The sequence of random variables $\sqrt{k}(\tilde{\theta}_{\text{ML}}(\mathbf{x}) - \theta_0)$ converge in distribution to normal distribution of variance $1/F(\theta_0)$ for any value of θ_0 :

$$\sqrt{k}(\tilde{\theta}_{\text{ML}}(\mathbf{x}) - \theta_0) \xrightarrow{d} \mathcal{N}(0, F(\theta_0)^{-1}). \quad (3.16)$$

Proof. (based on [Newey and McFadden [1994], Gundersen [2019]])

Let us introduce the normalized log-likelihood function:

$$L_k(\theta|\mathbf{x}) = \frac{1}{k} \log p(\mathbf{x}|\theta), \quad (3.17)$$

with $\mathbf{x} = \{x_1, \dots, x_k\}$. From the definition of the ML estimator

$$\tilde{\theta}_{\text{ML}}(\mathbf{x}) = \arg \max_{\theta} p(\{x_1, \dots, x_k\}|\theta) \Rightarrow \forall_{\mathbf{x}} \dot{L}_k(\tilde{\theta}_{\text{ML}}(\mathbf{x})|\mathbf{x}) = 0. \quad (3.18)$$

where sign dot denotes derivative $\frac{d}{d\theta}$. Assume without a loss that $\tilde{\theta}_{\text{ML}}(\mathbf{x}) \geq \theta_0$ (analogous reasoning may be performed for the opposite inequality). From mean value theorem there exists $\bar{\theta}(\mathbf{x})$ laying between θ_0 and $\tilde{\theta}_{\text{ML}}(\mathbf{x})$ such that:

$$\dot{L}_k(\tilde{\theta}_{\text{ML}}(\mathbf{x})|\mathbf{x}) = \dot{L}_k(\theta_0|\mathbf{x}) + \ddot{L}_k(\bar{\theta}(\mathbf{x})|\mathbf{x})(\tilde{\theta}_{\text{ML}}(\mathbf{x}) - \theta_0). \quad (3.19)$$

From Eq. (3.18) LHS is equal 0; after rearranging terms and multiplying both sides by $\frac{\sqrt{k}}{\ddot{L}(\bar{\theta}(\mathbf{x})|\mathbf{x})}$ we obtain:

$$\sqrt{k}(\tilde{\theta}_{\text{ML}}(\mathbf{x}) - \theta_0) = -\frac{\sqrt{k}\dot{L}_k(\theta_0|\mathbf{x})}{\ddot{L}(\bar{\theta}(\mathbf{x})|\mathbf{x})}. \quad (3.20)$$

Next, we will show the distribution convergence of the nominator and probability convergence of the denominator. Let us start with the second one.

For any θ from the weak law of large numbers (WLLN):

$$\ddot{L}_k(\theta|\mathbf{x}) = \frac{1}{k} \sum_{i=1}^k \frac{d^2}{d\theta^2} \log p(x_i|\theta) \xrightarrow{p} \mathbb{E} \left(\frac{d^2}{d\theta^2} \log p(x_1|\theta) \right) = -F(\theta), \quad (3.21)$$

where $\mathbb{E}(\cdot)$ denotes averaging. Next, as $\theta_0 \leq \bar{\theta}(\mathbf{x}) \leq \tilde{\theta}_{\text{ML}}(\mathbf{x})$ and $\tilde{\theta}_{\text{ML}}(\mathbf{x}) \xrightarrow{p} \theta_0$, also $\bar{\theta}(\mathbf{x}) \xrightarrow{p} \theta_0$, so combining with above (for sufficiently regular $L(\theta|\mathbf{x})$):

$$\ddot{L}_k(\bar{\theta}(\mathbf{x})|\mathbf{x}) \xrightarrow{p} -F(\theta_0). \quad (3.22)$$

The nominator is given as:

$$\sqrt{k} \dot{L}(\theta_0|\mathbf{x}) = \sqrt{k} \left(\frac{1}{k} \sum_{i=1}^k \frac{d}{d\theta} \log p(x_i|\theta) \Big|_{\theta=\theta_0} - \underbrace{\mathbb{E} \left(\frac{d}{d\theta} \log p(x_1|\theta) \right) \Big|_{\theta=\theta_0}}_{=0} \right) \quad (3.23)$$

therefore, from the central limit theorem, it converges to the normal distribution with variance:

$$\mathbb{E} \left(\left(\frac{d}{d\theta} \log p(x_1|\theta) \right) \Big|_{\theta=\theta_0}^2 \right) = F(\theta_0). \quad (3.24)$$

Applying above, together with Eq. (3.22), to Eq. (3.20) we finally get:

$$\forall_{\theta_0} \quad \sqrt{k}(\tilde{\theta}_{\text{ML}}(\mathbf{x}) - \theta_0) \xrightarrow{d} \mathcal{N} \left(0, \frac{F(\theta_0)}{F(\theta_0)^2} \right) = \mathcal{N} \left(0, \frac{1}{F(\theta_0)} \right), \quad (3.25)$$

what was to be proven. $\square$

3.4. Quantum Cramér-Rao bound

Previously we have asked how well the parameter θ value may be estimated by probing probability distribution $p(x|\theta)$, i.e., we have minimized MSE over all locally unbiased estimators. Here we generalize the problem slightly: there is a given family of states ρ_θ dependent on the unknown parameter θ , and we want to optimize the whole procedure over both choices of measurement $\{M_x\}$ as well as an estimator.

Using the result obtained for the classical case, we may see the problem as optimizing the classical FI Eq. (3.3) over the choice of measurement. After direct calculation, one would obtain [Braunstein and Caves [1994]]

$$\max_{\{M_x\}} F[p(x|\theta) = \text{Tr}(\rho_\theta M_x)] =: F_Q(\rho_\theta) = \text{Tr}(\rho_\theta L^2), \quad (3.26)$$

where L is symmetric logarithmic derivative (SLD) $\frac{d\rho_\theta}{d\theta} = \frac{1}{2} (L\rho_\theta + \rho_\theta L)$.

As for the tensor product of matrices $\rho_\theta \otimes \sigma_\theta$ corresponding SLD operators has the form $L_\rho \otimes \mathbb{1} + \mathbb{1} \otimes L_\sigma$, the QFI is additive:

$$F_Q(\rho_\theta \otimes \sigma_\theta) = F_Q(\rho_\theta) + F_Q(\sigma_\theta). \quad (3.27)$$

From the fact that QFI is FI maximized over measurements, it is also monotonic, i.e., for any θ -independent completely positive trace preserving map $\mathcal{E}(\cdot)$

$$F_Q(\rho_\theta) \geq F_Q(\mathcal{E}(\rho_\theta)), \quad (3.28)$$

as acting $\mathcal{E}(\cdot)$ followed by POVM on $\mathcal{E}(\rho_\theta)$ clearly defined equivalent POVM on ρ_θ directly.

Moreover, from above, it is also convex, i.e., for any θ -independent parameters c_ρ, c_σ

$$c_\rho F_Q(\rho_\theta) + c_\sigma F_Q(\rho_\sigma) \geq F_Q(c_\rho \rho_\theta + c_\sigma \rho_\sigma), \quad (3.29)$$

as one may consider the mapping $\mathcal{E} : c_\rho \rho_\theta \otimes |1\rangle\langle 1| + c_\sigma \rho_\sigma \otimes |2\rangle\langle 2| \mapsto c_\rho \rho_\theta + c_\sigma \rho_\sigma$. For the situation where c_ρ, c_σ depend on θ , the extended convexity may be defined [Alipour and Rezakhani [2015]], but it will not be discussed in this thesis.

As stated before, QFI may be derived by direct maximization of classical FI over the measurement. Here I want to present an alternative derivation based on joint optimization over both measurement and estimator (such an approach will be beneficial in multiparameter case Ch. 5). From the saturability of classical CR bound, we know that:

$$F_Q^{-1} = \min_{M_x, \text{l.u.}\tilde{\theta}(x)} \Delta^2 \tilde{\theta}. \quad (3.30)$$

From the discussion performed in Sec. 2.3, looking for the optimal strategy, we may restrict to protective measurements. For given θ_0 let me define:

$$X = \sum_x (\tilde{\theta}(x) - \theta_0) |x\rangle\langle x|. \quad (3.31)$$

Therefore QFI is given as:

$$F_Q^{-1} = \min_{M_x, \text{l.u.}\tilde{\theta}(x)} \Delta^2 \tilde{\theta} = \min_X [\text{Tr}(\rho_\theta X^2) \quad \text{with} \quad \text{Tr}(\dot{\rho} X) = 1], \quad (3.32)$$

where the minimization is made over all hermitian X . It may be performed with the Lagrange multipliers' method, from which:

$$\forall_{\delta X} \text{Tr}((\rho X + X \rho - \lambda \dot{\rho}) \delta X) = 0 \Rightarrow \dot{\rho} = \lambda(\rho X + X \rho), \quad (3.33)$$

where the last equality is equivalent to the definition of SLD (up to constant factor λ). Local

unbiasedness condition implies $\lambda = \text{Tr}(\rho_\theta L^2)$, and therefore

$$\text{Tr}(\rho_\theta X^2) = \frac{\text{Tr}(\rho_\theta L^2)}{(\text{Tr}(\rho_\theta L^2))^2} = \frac{1}{\text{Tr}(\rho_\theta L^2)} \Rightarrow F_Q = \text{Tr}(\rho_\theta L^2). \quad (3.34)$$

Here we should stress that even if we have derived the value of QFI with only local unbiasedness condition, in the limit of many repetitions, it has similar consequences as the classical one. Indeed, the corresponding measurement, i.e., the projection onto eigenvectors of L , results in the probability distribution maximizing classical FI. That indicates that this measurement will be optimal also from the global perspective and allow for effective usage of the ML estimator.

One issue should be mentioned here. In general, the measurement maximizing classical FI may depend on the exact value of the parameter, so the above reasoning cannot be applied directly. Still, in such case, one may asymptotically saturate quantum CR globally, using a two-step procedure [Gill and Massar [2000], Yang et al. [2019]]. First, use $\sim \sqrt{k}$ copies performing full tomography of the state [Haah et al. [2017]] to obtain some approximate value of θ_{est} with exponentially decreasing the risk of gross error $\propto e^{-\sqrt{k}}$. Then use the remaining $\sqrt{k}$ copies to perform the measurement in the basis of L for $\theta = \theta_{\text{est}}$. Consequently

$$\forall_{\theta \in \Theta} \lim_{k \rightarrow \infty} k \Delta^2 \tilde{\theta} = \frac{1}{F_Q(\rho_\theta)} \quad (3.35)$$

It is worth stressing that for pure states, the formula for QFI takes an especially simple form:

$$F_Q = 4(\langle \dot{\psi}_\theta | \dot{\psi}_\theta \rangle - |\langle \dot{\psi}_\theta | \psi_\theta \rangle|^2). \quad (3.36)$$

3.5. Quantum metrology – unitary evolution

Now we can move to a more general case, i.e., the problem of channel estimation, where the procedure may be optimized not only over measurement and estimator but also the input state or even additional features, as discussed in Sec. 2.5. For simplicity, let us start with the problem of unitary evolution, i.e., the one given as $U = e^{i\theta\Lambda}$. From the convexity of the QFI, for any channel, the optimal input state will be the pure state. Therefore, for a single usage of the channel, from Eq. (3.37), we have:

$$F_Q = 4(\langle \psi_\theta | \Lambda^2 | \psi_\theta \rangle - |\langle \psi_\theta | \Lambda | \psi_\theta \rangle|^2) = 4(\langle \psi | \Lambda^2 | \psi \rangle - |\langle \psi | \Lambda | \psi \rangle|^2) = 4\Delta^2 \Lambda. \quad (3.37)$$

In the case when generator Λ has a bounded spectrum with minimal and maximal eigenvalues λ_- , λ_+ , the state maximizing QFI will be an equal superposition of corresponding eigenstates $|\psi_\theta\rangle =$

$\frac{1}{\sqrt{2}}(e^{i\lambda-\theta}|\lambda_-\rangle + e^{i\lambda+\theta}|\lambda_+\rangle)$, for which

$$F_Q = (\lambda_+ - \lambda_-)^2. \quad (3.38)$$

The reasoning does not change significantly for n parallel usage of the gate Fig. 2.3(b). We simply need to focus on the states maximizing $\Delta^2 \Lambda^{\otimes n}$, which be the corresponding $n00n$ state $|\psi_\theta^n\rangle = \frac{1}{\sqrt{2}}(e^{in\lambda-\theta}|\lambda_-\rangle^{\otimes n} + e^{i\lambda+\theta}|\lambda_+\rangle^{\otimes n})$ with:

$$F_Q = n^2(\lambda_+ - \lambda_-)^2. \quad (3.39)$$

Finally, one may ask if applying a more advanced adaptive strategy Fig. 2.3(c) improves this result. The answer is no, which was shown [Giovannetti et al. [2006]] in the following way. For general adaptive evolution with n usage of gates, the output state is given by:

$$W_\theta|\psi\rangle, \quad \text{with} \quad W_\theta = V_n(U_\theta \otimes \mathbb{1})V_{n-1}(U_\theta \otimes \mathbb{1}) \cdots V_1(U_\theta \otimes \mathbb{1}). \quad (3.40)$$

so the FQI is equal

$$F_Q = 4(\langle \dot{\psi}_\theta | \dot{\psi}_\theta \rangle - |\langle \psi_\theta | \dot{\psi}_\theta \rangle|^2) = 4 \left(\langle \psi | \dot{W}_\theta^\dagger \dot{W}_\theta | \psi \rangle - |\langle \psi | W_\theta^\dagger \dot{W}_\theta | \psi \rangle|^2 \right) = 4\Delta^2 \left[-iW_\theta^\dagger \dot{W}_\theta \right], \quad (3.41)$$

where in the last step, identity $\mathbb{1} = W_\theta W_\theta^\dagger$ was added inside the first bracket, and factor $-i$ has been added to make the connection with the original evolution generator more visible. Indeed, $-iW_\theta^\dagger \dot{W}_\theta$ is simply sum of n properly rotated generators Λ

$$\begin{aligned} -iW_\theta^\dagger \dot{W}_\theta &= \sum_{j=1}^n \Lambda'_j(\theta), \quad \text{where} \\ \Lambda'_j(\theta) &= (V_{j-1}(U_\theta \otimes \mathbb{1}) \cdots V_1(U_\theta \otimes \mathbb{1}))^\dagger (\Lambda \otimes \mathbb{1}) (V_{j-1}(U_\theta \otimes \mathbb{1}) \cdots V_1(U_\theta \otimes \mathbb{1})), \end{aligned} \quad (3.42)$$

so its both minimal and maximal eigenvalues may be at most $n\lambda_-$, $n\lambda_+$, which is obtained with $\forall_j V_j = \mathbb{1}$ (as U_θ commutes with Λ). It is also clear that no additional entanglement with ancilla is needed. Therefore, optimal state may be simply obtain by n acting of the gates on single system copy: $|\psi_\theta^n\rangle = \frac{1}{\sqrt{2}}(e^{in\lambda-\theta}|\lambda_-\rangle + e^{i\lambda+\theta}|\lambda_+\rangle)$.

Note that states optimal in both parallel and adaptive schemes has the same problem as the one discussed in Sec. 2.1 – they became identical for parameters differ by $\frac{k2\pi}{n(\lambda_+-\lambda_-)}$ (up to the irrelevant global phase).

When the experimentalist has a large number of trials k to perform, it does not lead to a problem, as one may easily repeat the reasoning from the end of the previous section – first, perform the general tomography (for example, by using simple $|\psi_\theta\rangle = \frac{1}{\sqrt{2}}(e^{i\lambda-\theta}|\lambda_-\rangle + e^{i\lambda+\theta}|\lambda_+\rangle)$ state (with a

single action of the gate in each trial), and only after that focus on optimal local estimation. That would finally lead to:

$$\lim_{k \rightarrow \infty} k \Delta^2 \tilde{\theta} = \frac{1}{n^2(\lambda_+ - \lambda_-)^2}. \quad (3.43)$$

Still, more is needed to answer the general problem of **optimal usage of all N resources**. Intuitively one can see that starting from finite size Θ , to use all resources with at least close to optimal way, one could divide this N between different $n00n$ states (with different n), each devoted to estimating θ with a precision of another order. It was shown formally [Higgins et al. [2009]], that such an approach (assuming that the resources have been divided up in such a way as to allow sufficient repetition of the different orders of precision), indeed allows for obtaining quadratic scaling of the precision with N for the problem of phase shift in the interferometer:

$$\Delta^2 \tilde{\theta} \sim \frac{1}{N^2}. \quad (3.44)$$

However, a qualitative discussion of these results demands tools beyond QFI so that it will be performed later in [Ch. 4](#).

3.6. Infinite QFI for average energy constraints

In [Sec. 2.4](#), I have discussed the problem of the optimization of metrology protocol with constraint only on the average (not total) value of the energy, pointing at the fact that in such case $\Delta^2 H$ may be arbitrarily large. Let me go back here to this problem. Note that the above comment, together with [Eq. \(3.37\)](#), implies the possibility of obtaining the infinite value of QFI for finite average energy. How should it be interpreted in the context of the asymptotical saturability of CR? Let me discuss a simple example.

Let us consider the interferometer, where the input state is the mixture of $n00n$ states with different n , where the weight of the next state decreases like the inverse of the third power of n :

$$\rho = \frac{1}{\zeta(3)} \sum_{m=1}^{+\infty} \frac{1}{m^3} |m00m\rangle \langle m00m|, \quad (3.45)$$

where $\zeta(3) \approx 1.20$ is normalization factor. The mean energy is finite:

$$\langle H \rangle = \frac{1}{\zeta(3)} \sum_{m=1}^{+\infty} \frac{1}{2m^2} = \frac{\pi^2}{12\zeta(3)} \approx 0.64, \quad (3.46)$$

while its variance goes to infinity:

$$\langle \Delta^2 H \rangle = \frac{1}{\zeta(3)} \sum_{m=1}^{+\infty} \frac{1}{4m} = +\infty, \quad (3.47)$$

so does QFI. Knowing the measurement optimal for each $n00n$ state, consider the general measurement:

$$\{M_{m,\pm}\} = \{\frac{1}{2}(|0\rangle^{\otimes m} \pm |1\rangle^{\otimes m})(\langle 0|^{\otimes m} \pm \langle 1|^{\otimes m})\}, \quad (3.48)$$

for which $p_{m,+}(\theta) = \frac{1}{\zeta(3)m^3} \cos^2(m\theta/2)$, $p_{m,-}(\theta) = \frac{1}{\zeta(3)m^3} \sin^2(m\theta/2)$. This leads also to infinite classical FI. Moreover, such a formulated example is free from the problems found for a single $n00n$ state, discussed in [Sec. 2.1](#), i.e., the probability distribution distinct between any two values of θ . All these imply that for the ML estimator from equation [Eq. \(3.9\)](#), we have:

$$\lim_{k \rightarrow \infty} k \Delta^2 \tilde{\theta}_{\text{ML}} = 0. \quad (3.49)$$

How such a result should be understood? To understand it better, let us cut the sum in [Eq. \(3.45\)](#) at some large M to get:

$$\rho_M = \frac{1}{\mathcal{N}_M} \sum_{m=1}^M \frac{1}{m^3} |m00m\rangle \langle m00m|, \quad (3.50)$$

where $\mathcal{N}_M = \sum_{m=1}^M 1/m^3 \leq \zeta(3) \approx 1.20$ is normalization factor. For any finite M :

$$p_{m,+}(\theta) = \frac{\cos^2(m\theta/2)}{\mathcal{N}_M m^3}, \quad p_{m,-}(\theta) = \frac{\sin^2(m\theta/2)}{\mathcal{N}_M m^3} \implies F(M) = \left(\sum_{m=1}^M m^2/m^3 \right) / \mathcal{N}_M, \quad (3.51)$$

so classical FI is an unlimited growing function of M .

The MSE of ML estimator for different M has been plotted in figure [Fig. 3.2](#). The higher M , the bigger number of repetitions k is needed for saturating CR bound. Especially before saturating CR, we observe the scaling of MSE faster than $1/k$. Therefore for $M = +\infty$, we expect such scaling for all k , which is consistent with [Eq. \(3.49\)](#).

As the average energy used in a single trial is finite and equal ≈ 0.64 , the above indicates that for such a model, the final MSE will decrease faster than the inverse of the total energy used $\approx 0.64k$. What exactly will this scaling be? The answer for this equation is far from the above analysis obtained by QFI. However, from [Eq. \(2.12\)](#), we can immediately say that it cannot decrease faster than quadratically with total average energy used $\sim \frac{1}{k^2}$.

3.7. General theorem for a noisy estimation

As stated in the introduction, the main object of interest in this thesis will be the unitary noiseless evolution, therefore from a strictly theoretical point of view, the discussion performed so far is enough for understanding the context of the results presented later in [Ch. 4](#). However, this is the idealized situation, as in practice, some noise occurs in any actual metrological situation. Therefore, for completeness, I present the state-of-the-art general theorem about the possibility of attaining the Heisenberg scaling in noisy metrology, formulated only within QFI formalism. This presentation

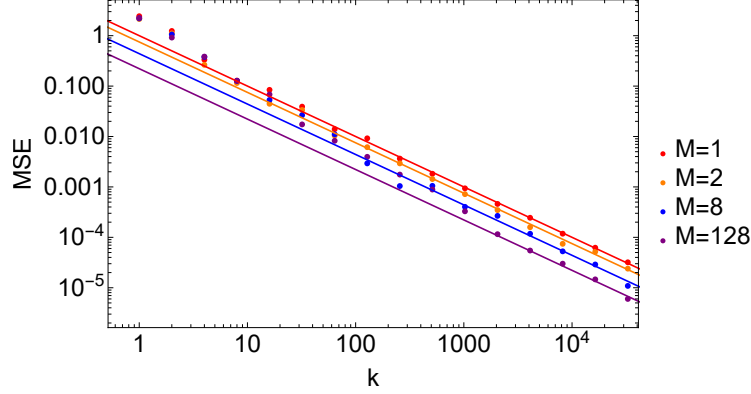


Figure 3.2: MSE of ML estimator vs. trials number k for the model Eq. (3.51) is plotted. Results are presented for $M = 1, 2, 8, 128$. The dots are the MSE drawn from $p_{m,+}(\theta), p_{m,-}(\theta)$ probabilities at $\theta = 1.78072$ (to reduce statistical noise, for each k , the MSE was calculated 100 and averaged). Solid lines are corresponding CR bounds.

is only to give a broader context to the results, and the noise case will not be analyzed quantitatively later in this work.

For physical system $\mathcal{H}_S$, consider general channel $\mathcal{E}_\theta : \mathcal{H}_S \rightarrow \mathcal{H}_S$, i.e., completely positive trace preserving (CPTP) map. Any such map may be defined by the set of (not unique) Kraus operators $\{K_k(\theta)\}$

$$\mathcal{E}_\theta(\rho) = \sum_{k=0}^r K_k(\theta) \rho K_k^\dagger(\theta), \quad (3.52)$$

where $\{K_k(\theta)\}_k$ satisfies $\sum_{i=1}^r K_i(\theta)^\dagger K_i(\theta) = \mathbb{1}$ (an alternative way of characterizing quantum channels via Choi matrices will be given in Sec. 5.3). Further, for more compact notation, the parameter in the bracket will be omitted $K_k := K_k(\theta)$, but one should keep in mind that they depend on θ .

Then the QFI of the output state maximized over the input state is bounded by [Fujiwara and Imai [2008], Demkowicz-Dobrzański et al. [2012]]

$$\max_{\rho} F_Q \leq 4 \left\| \sum_{k=0}^r \dot{K}_k^\dagger \dot{K}_k \right\|, \quad (3.53)$$

which comes from the fact that QFI of any mixed state ρ_θ is smaller or equal to QFI of its purification (as the consequence of Eq. (3.28)). Next, as the set of Kraus operators is defined only up to unitary transformation: [Fujiwara and Imai, 2008, Proposition 2.]

$$\tilde{K}_k(\theta) = \sum_j u_{kj}(\theta) K_j(\theta), \quad (3.54)$$

(where in principle the unitary $u_{kj}(\theta)$ mixing different Krauss operator may also depend on θ), the inequality may be tightened by performing minimization over all Krauss representation. Around θ_0 , denoting $u_{kj}(\theta) = u_{kj}(\theta_0) + ih_{kj}(\theta - \theta_0) + \mathcal{O}((\theta - \theta_0)^2)$ (where h is hermitian), by direct subjection we get:

$$\max_{\rho} F_Q \leq \min_h 4 \left\| \sum_{k=0}^r \dot{K}_k^\dagger \dot{K}_k \right\| = \min_h 4 \left\| \sum_{k=0}^r \left(\dot{K}_k^\dagger + i \sum_j h_{kj} K_j^\dagger \right) \left(\dot{K}_k - i \sum_j h_{kj} K_j \right) \right\|, \quad (3.55)$$

Moreover, the above bound is proved to be saturable if ancilla allowed $\rho \in \mathcal{H}_S \otimes \mathcal{H}_A$. Starting from this bound, one may derive practical bounds for parallel and adaptive strategies discussed in Fig. 2.3(b,c) using n quantum gates, in terms of the Krauss operators of a single gate.

For the parallel scheme [Fujiwara and Imai [2008]]:

$$F_Q \leq \min_h 4 (n\|\alpha\| + n(n-1)\|\beta\|^2) \quad (3.56)$$

where

$$\begin{aligned} \alpha &= \sum_{k=0}^r \dot{K}_k^\dagger \dot{K}_k = \sum_{k=0}^r \left(\dot{K}_k^\dagger + i \sum_j h_{kj} K_j^\dagger \right) \left(\dot{K}_k - i \sum_j h_{kj} K_j \right), \\ \beta &= \sum_{k=0}^r \tilde{K}_k^\dagger \dot{K}_k = \sum_{k=0}^r K_k^\dagger \left(\dot{K}_k - i \sum_j h_{kj} K_j \right). \end{aligned} \quad (3.57)$$

Note especially that $\|\alpha\| \geq \|\beta\|^2$.

For a general sequential adaptive scheme, the bound may be formulated in the iterative form [Kurdzialek et al. [2022]] (which is an improved version of the bound form [Demkowicz-Dobrzański and Maccone [2014]]):

$$F_Q \leq 4a^{(n)}, \quad \text{where} \quad a^{(i+1)} = \min_h \left[a^{(i)} + \|\alpha\| + 2\|\beta\|\sqrt{a^{(i)}} \right], \quad \text{with} \quad a^{(0)} = 0, \quad (3.58)$$

which may be further bounded by two alternative closed formulas:

$$F_Q \leq \min_h \begin{cases} 4 \left(n\|\alpha\| + n(n-1)\|\beta\|\sqrt{\|\alpha\|} \right), \\ 4 \left(n\|\alpha\| + n(n-1)\|\beta\|^2 + (\|\alpha\| - \|\beta\|^2)n \log n \right), \end{cases} \quad (3.59)$$

where the first one is more usefull in regime $n \approx 1$, $\|\beta\|/\sqrt{\|\alpha\|} \ll 1$, while the second one in the regime $n \gg 1$, $\|\beta\|/\sqrt{\|\alpha\|} \approx 1$.

Looking at both Eq. (3.56) and Eq. (3.59) one can see that the necessary condition for occurring HS

is that $\forall_h \|\beta\| \neq 0$. It has been identified as a "Hamiltonian-not-in-Kraus-span" (HNKS) condition, namely for:

$$\tilde{\Lambda} = i \sum_k K_k^\dagger \dot{K}_k, \quad \mathcal{S} = \text{span}_{\mathbb{H}}\{K_i^\dagger K_j \text{ for all } i, j\} \quad (3.60)$$

this condition is satisfied iff $\tilde{\Lambda} \notin \mathcal{S}$. Otherwise, in the limit of large n , both bounds reduce to:

$$F_Q \lesssim \min_{h:\beta=0} 4n\|\alpha\|, \quad (3.61)$$

or, more formally, $\lim_{n \rightarrow \infty} \frac{1}{n} F_Q = \min_{h:\beta=0} \|\alpha\|$. In such case, the bound has been proven to be asymptotically tight [Zhou and Jiang, 2021, Theorem 2] :

$$\lim_{n \rightarrow \infty} n F_Q = 4\|\alpha\|. \quad (3.62)$$

The most prominent example of practical importance, when the above bound has been saturated, revealing quantum advantage, was using squeezed vacuum state [Caves [1981]] in the interferometer in the LIGO gravitational wave detector [LIGO [2011], Aasi et al. [2013]], where the experimentalists obtain the MSE improved by factor respectively 3.5dB and 2.15dB compared to the optimal "classical" strategy with the same output energy. The value of MSE is in perfect accord with theoretical model [Aasi et al. [2013]], see also App. A for exact formula and relation with the bound.

In the case where $\min_h \|\beta\| > 0$, in the limit of large n the bounds Eq. (3.56), Eq. (3.59) (the second line in the latter one) reduce to:

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} F_Q \leq \min_h 4\|\beta\|^2 \quad (3.63)$$

and it may be saturated [Zhou and Jiang [2021]], up to this leading term. The authors of [Zhou and Jiang [2021]] propose the following procedure.

They show that there exists two dimensional subspace of joint system and ancilla $\text{span}\{|0_L\rangle, |1_L\rangle\} \in \mathcal{H}_S \otimes \mathcal{H}_A$ and the recovery operation $\mathcal{R} : \text{Lin}(\mathcal{H}_S \otimes \mathcal{H}_A) \rightarrow \text{Lin}(\text{span}\{|0_L\rangle, |1_L\rangle\})$ such that for any $\rho \in \text{Lin}(\text{span}\{|0_L\rangle, |1_L\rangle\})$ in the closest neighborhood of θ_0 , effectively unitary evolution rotating by the angle $2\|\beta\|(\theta - \theta_0)$ is obtained:

$$\mathcal{R}(\mathcal{E}_\theta \otimes \mathbb{1}_A(\rho)) \stackrel{\theta \approx \theta_0}{\approx} e^{i\|\beta\|(\theta - \theta_0)\sigma_z} \rho e^{-i\|\beta\|(\theta - \theta_0)\sigma_z}. \quad (3.64)$$

More precisely, exactly at point θ_0 this procedure satisfies¹:

$$\begin{aligned}\mathcal{R}(\mathcal{E}_{\theta_0} \otimes \mathbb{1}_A(\rho)) &= \rho \\ \frac{d}{d\theta} \mathcal{R}(\mathcal{E}_{\theta} \otimes \mathbb{1}_A(\rho)) \Big|_{\theta=\theta_0} &= i\|\beta\|(\sigma_z \rho - \rho \sigma_z)\end{aligned}\tag{3.65}$$

To get some intuition about the applicability of this protocol, let us consider the simplest non-trivial example.

Example

Consider a qubit sensing the phase-signal in z direction, subjected to dephasing noise in x direction. What is important, the dephasing operator act before the unitary operator. The exemplary Kraus representation for this model is:

$$K_1 = p e^{i\theta \sigma_z}, \quad K_2 = \sqrt{1-p} e^{i\theta \sigma_z} \sigma_x \tag{3.66}$$

for which:

$$\alpha = \mathbb{1} \Rightarrow \|\alpha\| = 1, \tag{3.67}$$

$$\beta = -i(p\sigma_z + (1-p)\sigma_x\sigma_z\sigma_x) = -i(1-2p)\sigma_z \Rightarrow \|\beta\| = |1-2p|. \tag{3.68}$$

Instead of showing that this representation minimizes $\|\beta\|$, we just show the exemplary protocol saturating Eq. (3.63) (so optimal $h = 0$),

Let $|0_L\rangle = |00\rangle \in \mathcal{H}_S \otimes \mathcal{H}_A$ and $|1_L\rangle = |11\rangle \in \mathcal{H}_S \otimes \mathcal{H}_A$. Then for the state $|\psi\rangle = a|0_L\rangle + b|1_L\rangle = |\psi\rangle = a|00\rangle + b|11\rangle$ we have:

$$\begin{aligned}\mathcal{E}_{\theta}(|\psi\rangle\langle\psi|) &= \\ & p(ae^{-i\theta}|00\rangle + be^{+i\theta}|11\rangle)(ae^{+i\theta}\langle 00| + be^{-i\theta}\langle 11|) \\ & + (1-p)(ae^{+i\theta}|10\rangle + be^{-i\theta}|01\rangle)(ae^{-i\theta}\langle 10| + be^{+i\theta}\langle 01|)\end{aligned}\tag{3.69}$$

Consider recovery channel $\mathcal{R}$ defined by Krauses R_k :

$$R_1 = e^{+i\theta_0}|00\rangle\langle 00| + e^{-i\theta_0}|11\rangle\langle 11|, \quad R_2 = e^{-i\theta_0}|00\rangle\langle 10| + e^{+i\theta_0}|11\rangle\langle 01| \tag{3.70}$$

(as $\mathcal{R}$ is CPTP map, it may be simulated by acting of unitaries, if sufficiently large ancilla allowed so that it may be seen as the special case of Fig. 2.3(c)). Then (restricted to subspace $|0_L\rangle, |1_L\rangle$):

$$\mathcal{R}(\mathcal{E}_{\theta}(|\psi\rangle\langle\psi|)) = \begin{bmatrix} |a|^2 & ab^* (pe^{-i(\theta-\theta_0)} + (1-p)e^{+i(\theta-\theta_0)}) \\ a^*b (pe^{+i(\theta-\theta_0)} + (1-p)e^{-i(\theta-\theta_0)}) & |b|^2 \end{bmatrix} \tag{3.71}$$

¹in original paper [Zhou and Jiang 2021] the additional mapping between 2-dimensional logical space to physical space $\text{span}\{|0_L\rangle, |1_L\rangle\} \in \mathcal{H}_S \otimes \mathcal{H}_A$ appears, which I omitted for simplicity of the formulas

Exactly at point θ_0 above satisfies [Eq. \(3.65\)](#), and the bound [Eq. \(3.63\)](#) may be obtain by choosing as the input state $|\psi\rangle = \frac{1}{\sqrt{2}}(|0_L\rangle^{\otimes n} + |1_L\rangle^{\otimes n})$. Note, however, that for finite difference $\theta - \theta_0$ the factor multiplying off-diagonal elements has the absolute value $|pe^{-i(\theta-\theta_0)} + (1-p)e^{+i(\theta-\theta_0)}| = \sqrt{1 - 4p(1-p)\sin^2(\theta - \theta_0)}$, so to make it negligible in analyzing the evolution of the state of the form $|\psi\rangle$ it is necessary $\sqrt{1 - 4p^2(1-p^2)\sin^2(\theta - \theta_0)}^n \approx 1 \Rightarrow (\theta - \theta_0) \ll \frac{1}{\sqrt{n}}$.

Again, this is not a problem when the number of gates that can be used in a single implementation n is limited, while the experiment may be repeated a large number of times k . For such case, one may obtain:

$$\lim_{k \rightarrow \infty} k\Delta^2\tilde{\theta} = \frac{1}{n^2\|\beta\|^2}. \quad (3.72)$$

However, the possibility of using such a protocol in a situation where one wants to use whole N resources optimally in analogous to the procedure discussed at the and of [Sec. 3.5](#) is highly nontrivial. Namely, it is not clear if, for sequential usage of $n00n$ state with increasing n , up-to-date knowledge of θ will be sufficient for effectively applying QEC procedure in the next step in such a way to provide finally

$$\Delta^2\tilde{\theta} \stackrel{?}{\sim} \frac{1}{N^2}, \quad (3.73)$$

in analogous to [Eq. \(3.44\)](#), proven for unitary noisless estimation in [\[Higgins et al. \[2009\]\]](#). This, however, remain an open question outside of the topic of this thesis.

CHAPTER 4

OBTAINABLE HEISENBERG LIMIT IN SINGLE PARAMETER UNITARY ESTIMATION

In this chapter, I will discuss the fundamental achievable bounds for the precision obtainable in single parameter unitary estimation with a total amount of resources limited. As mentioned before, this problem cannot be analyzed with the usage of QFI or any other formalism based solely on the concept of the local unbiasedness of the estimator. Therefore to perform quantitative analysis, I will use the formalism of minimax estimation and Bayesian estimation.

4.1. Bayesian and minimax costs

Let me start again with the "classical" theory of estimation, i.e., where the probability distribution of the measure outcomes dependent on the value of the parameter $p(x|\theta)$ is fixed.

The most popular alternative for FI formalism is the Bayesian one. It postulates that the unknown parameter θ may take a given value according to some known a priori distribution $p(\theta)$. In this sense, both θ and x are treated as random variables with joint probability distribution $p(x, \theta) = p(x|\theta)p(\theta)$. That implies marginal distribution $p(x) = \int d\theta p(x|\theta)p(\theta)$, and also conditional probability:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int d\theta p(x|\theta)p(\theta)}, \quad (4.1)$$

also called a posteriori distribution of θ after obtaining the result x . By Bayesian MSE, we define the standard MSE, averaged with a priori distribution:

$$\overline{\Delta^2 \tilde{\theta}} = \int p(\theta) d\theta \int dx p(x|\theta) (\tilde{\theta}(x) - \theta)^2. \quad (4.2)$$

This approach may be applied to the problem, when the parameter to be measured truly fluctuate according to known distribution (as, for example, the temperature in a given room); however, it may be also applied in measuring the parameter which is postulated to be constant (like, for example, transition frequency of an atom). In the second case, the a priori probability $p(\theta)$ will express our imperfect knowledge, coming from previous measurements.

The alternative for the Bayesian approach is the minimax one [Hájek [1972], Hayashi [2011]]. Here one only needs to assume that the actual value of the parameter belongs to some set $\theta \in \Theta$, and then to validate the effectiveness of the estimation strategy, one assumes the most pessimistic scenario, i.e., maximizes the MSE over all possible values of parameter:

$$\widehat{\Delta^2 \tilde{\theta}} = \sup_{\theta \in \Theta} \int dx p(x|\theta) (\tilde{\theta}(x) - \theta)^2. \quad (4.3)$$

Directly from the definition, one can see that if a priori distribution in the Bayesian approach has support inside of the set considered in the minimax approach $\text{supp } p(\theta) \subseteq \Theta$, then for any fixed estimator, the Bayesian cost is smaller or equal than the minimax one:

$$\forall_{\tilde{\theta}(\cdot)} \quad \overline{\Delta^2 \tilde{\theta}} \leq \widehat{\Delta^2 \tilde{\theta}}. \quad (4.4)$$

Bayesian cost is also connected with FI, via the van Trees inequality [Gill and Levit [1995]]:

$$\overline{\Delta^2 \tilde{\theta}} \geq \frac{1}{\int d\theta p(\theta) F(\theta) + I}, \quad I = \int d\theta \frac{1}{p(\theta)} \left(\frac{dp(\theta)}{d\theta} \right)^2, \quad (4.5)$$

where I should be understood as the information included in a priori distribution.

From the asymptotic optimality of the ML estimator, in the limit of many repetitions:

$$\min_{\tilde{\theta}} \lim_{k \rightarrow \infty} k \overline{\Delta^2 \tilde{\theta}} = \lim_{k \rightarrow \infty} k \overline{\Delta^2 \tilde{\theta}}_{\text{ML}} = \int p(\theta) d\theta \frac{1}{F(\theta)}, \quad (4.6)$$

$$\min_{\tilde{\theta}} \lim_{k \rightarrow \infty} k \widehat{\Delta^2 \tilde{\theta}} = \lim_{k \rightarrow \infty} k \widehat{\Delta^2 \tilde{\theta}}_{\text{ML}} = \sup_{\theta \in \Theta} \frac{1}{F(\theta)}. \quad (4.7)$$

Especially these two coincide if $F(\theta)$ does not depend on θ .

Now we want to use these formalisms to analyze the problem of quantum metrology, i.e., $p(x|\theta) = \text{Tr}(M_x \rho_\theta)$. However, before starting, two issues should be mentioned here.

First, we focus on the situation where all resources are used in the optimal way, so, as discussed before, all resources are accumulated in a single measurement realization. This implies that any single measurement outcome is connected directly with the estimator indication (unlike in the previous case, where the estimator could be the function of a sequence of outcomes). Therefore, to simplify notation, we may label the measurement directly by the value of estimator $(\tilde{\theta}(x) - \theta)^2 \text{Tr}(M_x \rho_\theta) dx \rightarrow (\tilde{\theta} - \theta)^2 \text{Tr}(\rho_\theta M_{\tilde{\theta}}) d\tilde{\theta}$.

Second, as we are no longer restricted only to local estimation, we would like to consider a more general cost than the quadratic one. Therefore, in general, we will want to minimize the mean value of cost function $\mathcal{C}(\theta, \tilde{\theta})$, which we typically choose to satisfy $\mathcal{C}(\theta, \tilde{\theta}) \stackrel{|\tilde{\theta}-\theta| \ll 1}{\approx} (\tilde{\theta} - \theta)^2$, which allows us to maintain a relationship with the results obtained using Fisher information. The proper formulas in both formalisms are, respectively, in Bayesian formalism:

$$\bar{\mathcal{C}} = \int p(\theta) d\theta \int d\tilde{\theta} \text{Tr}(\rho_\theta M_{\tilde{\theta}}) \mathcal{C}(\theta, \tilde{\theta}), \quad (4.8)$$

and in minimax formalism:

$$\widehat{\mathcal{C}} = \sup_{\theta \in \Theta} \int d\tilde{\theta} \text{Tr}(\rho_{\theta} M_{\tilde{\theta}}) \mathcal{C}(\theta, \tilde{\theta}). \quad (4.9)$$

4.2. Optimality of covariant measurements

Before discussing the fundamental bound derived in [Górecki et al. [2020]], I want to recall the theorem, which significantly simplifies the measurement optimization when the estimation problem satisfies proper symmetry conditions. While this theorem was not used in [Górecki et al. [2020]], and it is not necessary for understanding it, it allows for deriving very expressive examples providing the reader with good intuition.

Consider a problem where the family of channels is a unitary representation of a compact Lie group G , such that $\mathcal{E}_g(\rho) = U_g \rho U_g^\dagger$ and the aim is to estimate the group element $g \in G$. We would say that the estimation problem is covariant (in both Bayesian or minimax approach), iff:

- the cost function is invariant under the acting of the group $\mathcal{C}(g_1, g_2) = \mathcal{C}(hg_1, hg_2)$.
- (required in Bayesian formalism) a priori distribution is invariant under acting of the group $p(g)dg = p(hg)d(hg)$ – the prior is uniform with respect to the Haar measure on the group. For simplicity of the notation, further, I assume that dg is the normalized Haar measure, $\int_G dg = 1$ (so covariant a priori distribution is $p(g) = 1$).

Then for a given input state ρ with single usage of the channel, the optimal cost is obtainable with covariant measurement, i.e., the one of the form [Holevo, 1982, Chapter 4]:

$$M_g = U_g^\dagger M_e U_g, \quad \int dg U_g^\dagger M_e U_g = \mathbb{1}. \quad (4.10)$$

After applying it to Eq. (4.8), Eq. (4.9) it simplified the problem to the form:

$$\left. \begin{aligned} \inf_{M_{\tilde{g}}} \widehat{\mathcal{C}} &= \inf_{M_{\tilde{g}}} \sup_{\theta \in \Theta} \int d\tilde{g} \text{Tr}(U_g \rho U_g^\dagger M_{\tilde{g}}) \mathcal{C}(g, \tilde{g}) \\ \inf_{M_{\tilde{g}}} \overline{\mathcal{C}} &= \inf_{M_{\tilde{g}}} \int dg \int d\tilde{g} \text{Tr}(U_g \rho U_g^\dagger M_{\tilde{g}}) \mathcal{C}(g, \tilde{g}) \end{aligned} \right\} = \inf_{M_e} \int d\tilde{g} \text{Tr}(\rho U_{\tilde{g}}^\dagger M_e U_{\tilde{g}}) \mathcal{C}(e, \tilde{g}) \quad (4.11)$$

The proof is based on the observation that for any measurement $\{M'_g\}$ one may construct corresponding covariant measurement by proper averaging: $M_g := U_g \left(\int dh U_h^\dagger M'_h U_h \right) U_g^\dagger$, while this procedure does not affect the Bayesian cost and may only decrease the minimax cost. Moreover, within minimax formalism, the statement may be generalized also for non-compact groups [Bogomolov [1982]], where formal averaging may not be possible.

The above directly implies that the covariant measurement is optimal for any parallel strategy with

the usage of N gates – as $U_g^{\otimes N}$ is still a group representation. In [Ch. 5.3](#), I will reproduce the proof of an even stronger statement, saying that if ancilla is allowed, optimal parallel scheme with covariant measurement cannot be beaten by any adaptive strategy (the derivation, however, will demand more advanced formalism, which is not necessary at this point).

4.3. Interferometer with a fixed total number of photons

The most canonical example of applying the above theorem is the problem of estimating the unknown phase in the interferometer. In such case the bound [Eq. \(2.5\)](#) gives:

$$\Delta^2 \tilde{\theta} \geq \frac{1}{N^2}, \quad (4.12)$$

without guarantee of saturability. It was shown [[Luis and Peřina \[1996\]](#), [Buřek et al. \[1999\]](#), [Berry and Wiseman \[2000\]](#)], that for estimation of a completely unknown phase the additional π^2 factor appears.

First, note that the value θ should be identified with $\theta + 2\pi$; therefore, we need to choose a proper cost function considering this fact. The common choice is

$$\mathcal{C}(\theta, \tilde{\theta}) = 4 \sin^2 \left(\frac{\tilde{\theta} - \theta}{2} \right), \quad (4.13)$$

which for small difference $\tilde{\theta} - \theta$ may be well approximated by $(\tilde{\theta} - \theta)^2$. The average cost for the covariant measurement is therefore equal:

$$\bar{\mathcal{C}} = \frac{1}{2\pi} \int d\tilde{\theta} \text{Tr}(\rho_0 U_{\tilde{\theta}} M_0 U_{\tilde{\theta}}^\dagger) 4 \sin^2 \left(\frac{\tilde{\theta}}{2} \right), \quad (4.14)$$

(so in notation from [Eq. \(4.11\)](#) $g = \frac{\theta}{2\pi}$). Without loss, we may restrict ourselves to fully symmetric space and then use the eigenbasis of $U_{\tilde{\theta}}$. Since the mean cost is linear in ρ , we may restrict to a pure input state $|\psi^N\rangle$, for which

$$U_{\theta} |\psi^N\rangle = U_{\theta} \left(\sum_{m=0}^N c_m |m\rangle \right) = \sum_{m=0}^N e^{im\theta} c_m |m\rangle, \quad (4.15)$$

where $|m\rangle$ is the state with m photons in the upper (sensing) arm. After applying identity $4 \sin^2 \left(\frac{\tilde{\theta}}{2} \right) = 2 - e^{i\tilde{\theta}} - e^{-i\tilde{\theta}}$ we obtain

$$\bar{\mathcal{C}} = \int \frac{d\theta}{2\pi} \left(\sum_{jk} c_j^* c_k e^{ik\theta} M_{kj} e^{-ij\theta} \right) (2 - e^{i\theta} - e^{-i\theta}) = \sum_{jk} c_j^* c_k M_{kj} (2\delta_{j,k} - \delta_{j+1,k} - \delta_{j-1,k}). \quad (4.16)$$

From condition [Eq. \(4.10\)](#) $\forall_m M_{mm} = 1$; moreover, without loss of generality, we may assume that $\forall_m c_m \geq 0$ (otherwise, one can always redefine the vectors of the basis, by multiplying them by

proper phase factors). Therefore, the minimum value will be obtained for the maximal possible $M_{m,m+1}$. From positivity condition for M , $M_{m,m+1} \leq 1$, where the equality is satisfied if one chooses as M the projection onto vector $\sum_{m=0}^N |m\rangle$, such that:

$$\frac{1}{2\pi} M_{\tilde{\theta}} = |\tilde{\theta}\rangle\langle\tilde{\theta}|, \quad \text{with} \quad |\tilde{\theta}\rangle = \frac{1}{\sqrt{2\pi}} \sum_{m=0}^N e^{im\tilde{\theta}} |m\rangle. \quad (4.17)$$

After that, the problem simplifies to:

$$\min_{|\psi^N\rangle, \{M_{\tilde{\theta}}\}} \bar{\mathcal{C}} = \min_{\{c_m\}} \sum_{m=1}^{N-1} (2c_m c_m - c_m c_{m+1} - c_m c_{m-1}), \quad (4.18)$$

which is equivalent to finding the lowest eigenvector of the following tri-diagonal matrix:

$$\begin{bmatrix} 2 & -1 & 0 & 0 & & \\ -1 & 2 & -1 & 0 & \dots & \\ 0 & -1 & 2 & -1 & & \\ 0 & 0 & -1 & 2 & & \\ & \vdots & & & \ddots & -1 \\ & & & & -1 & 2 \end{bmatrix}. \quad (4.19)$$

The exact solution is known:

$$|\psi^N\rangle = \sqrt{\frac{1}{N+2}} \sum_{m=0}^N \sin\left(\frac{(m+1)\pi}{N+2}\right) |m\rangle \quad (4.20)$$

for which

$$\bar{\mathcal{C}} = 2 \left(1 - \cos\left(\frac{\pi}{N+2}\right) \right) = \frac{\pi^2}{N^2} + \mathcal{O}\left(\frac{1}{N^3}\right), \quad (4.21)$$

so indeed additional π^2 factor appears when compared with [Eq. \(4.12\)](#).

While the above provides the exact and complete solution to the problem, some comments are worth mentioning here.

First, while above we have used covariant measurement, the same cost may also be obtained by using the projective measurement, which may be seen as the discrete version of the previously discussed:

$$\{M_j\} = \{|\chi_j\rangle\langle\chi_j|\}_{j=0}^N, \quad |\chi_j\rangle = \frac{1}{\sqrt{N+1}} \sum_{m=0}^N e^{i\frac{2\pi j}{N+1}m} |m\rangle, \quad \langle\chi_j|\chi_k\rangle = \delta_{jk}. \quad (4.22)$$

(note that this is a specific feature of this peculiar model and does not need to be satisfied for

arbitrary covariant problems). This choice also has an intuitive interpretation, as this basis may be obtained by applying discrete Fourier transform to the basis with well-defined photons number in upper arm $\{|m\rangle\}_{m=0}^N$. Therefore, it stays in the analog to the problem of the position shift estimation in continuous space, where the shift generator is momentum operator $\hat{p} = \frac{1}{i} \frac{d}{dx}$, while the optimal measurement is $\{|x\rangle\}$ (which may be obtained by applying Fourier transform to $\{|p\rangle\}$). Nevertheless, adequacy is not exact, as for the phase θ (unlike for position x), no corresponding observable exists.

Second, even if the problem is analytically solvable for finite N , it will be useful to see what approximation may be used in the limit $N \rightarrow \infty$ (that will allow us to understand better a more general case). Let us introduce $m/N \rightarrow \mu$ and consider $\hat{f}(\mu)$ and the state Eq. (4.15) with $c_m = \frac{1}{\sqrt{N}} \hat{f}(m/N)$. Then Eq. (4.18) takes form:

$$\min_{|\psi^N\rangle, \{M_{\hat{\theta}}\}} \bar{\mathcal{C}} = \min_{\hat{f}} \frac{1}{N^2} \cdot \frac{1}{N} \sum_{m=1}^N \hat{f}(m/N) \frac{2\hat{f}(m/N) - \hat{f}(m/N + 1/N) - \hat{f}(m/N - 1/N)}{(1/N)^2}, \quad (4.23)$$

which, assuming that $\hat{f}(\mu)$ does not change very rapidly at distance of order $1/N$, may be approximated by (see [Imai and Hayashi [2009], Hayashi [2016]] for more discussion about exact correspondence in the limit $N \rightarrow \infty$):

$$\min_{|\psi^N\rangle, \{M_k\}} \bar{\mathcal{C}} \approx \frac{1}{N^2} \min_f \int_0^1 d\mu \hat{f}^*(\mu) \left(-\frac{d^2}{d\mu^2} \right) \hat{f}(\mu), \quad \hat{f}(0) = \hat{f}(1) = 0, \quad (4.24)$$

so the problem is equivalent to minimizing the kinetic energy of a single particle in the infinite well. The solution is then $\hat{f}(\mu) = \sqrt{2} \sin(\mu\pi)$, for which the cost is $\frac{\pi^2}{N^2}$.

Third, the vital difference between minimizing variance and minimizing the risk of coarse error in this context is worth noting. In the many repetitions scenario, in the limit of large k , for the ML estimator, the distribution of difference $\tilde{\theta}_{\text{ML}} - \theta$ converges to the gaussian one with the variance k , so the probability of making gross error decreases exponentially with the number of resources. Contrary, in the case discussed in this section, the probability of getting outcome $\tilde{\theta}$ where the actual value of the parameter is θ is given as (for optimal covariant measurement Eq. (4.17)):

$$\begin{aligned} p(\tilde{\theta}|\theta) &= \frac{1}{2\pi} \left(\sum_{k=0}^N e^{im\tilde{\theta}} e^{-im\theta} c_m \right)^2 \approx \frac{1}{2\pi} \left(\int_0^1 d\mu e^{iN\mu(\tilde{\theta}-\theta)} \hat{f}(\mu) \right)^2 = \\ &= |f(N(\tilde{\theta} - \theta))|^2 = \frac{2\pi(1 + \cos(N(\tilde{\theta} - \theta)))}{(N^2(\tilde{\theta} - \theta)^2 - \pi^2)^2}, \end{aligned} \quad (4.25)$$

so it decreases only like forth power of N . That is – the strategy using all resources to minimize the quadratic cost (or the one close to quadratic) turns out to be extremely poor for the problem

of discriminating the final sized region to which θ belongs. It was discussed in [Imai and Hayashi [2009]] that for optimal usage of all resources, these two problems need to be discussed independently, and completely different input states turn out to be optimal in both of them (while in the many repetitions scenario, the same strategy optimized both problems).

At last, the broad literature is dedicated to analyzing the relation between the results obtained for the phase estimation in both QFI and Bayesian formalism. The complex analysis of plenty of measurement strategies may be found in [Berry et al. [2009]]. As mentioned in Ch. 3, in [Higgins et al. [2009]], it was analytically proven that the Heisenberg scaling with all resources N may be obtained by a proper sequence of the measurement for $n00n$ states with different n in a different iteration, which was also demonstrated in the experiment [Higgins et al. [2007, 2009]]. In [Kaftal and Demkowicz-Dobrzański [2014]], the numerical analysis suggests that the factor π^2 may be exactly obtained for such strategy if one allows for collective measurement on the quantum superposition of all these $n00n$ states.

4.4. Interferometer with a fixed average number of photons

The same problem may be discussed with a constraint on the average number of the photons in the sensing arm, which is the constraint for the mean energy in understanding $\langle \psi | H | \psi \rangle = \bar{N}$, as discussed in Sec. 2.4. Here I recall the results from [Summy and Pegg [1990], Bandilla et al. [1991]].

The reasoning remains almost unchanged compared to the case with a fixed number of photons N . The only difference is that in equation Eq. (4.18), the sum is over all natural indices $\sum_{m=0}^{\infty}$ and the additional constraint appears $\sum_{m=0}^{\infty} \rho_{mm} m = \bar{N}$. Similarly, as in the previous case, by approximating discrete variables by continuous one for large $\bar{N}$ ($m/\bar{N} \rightarrow \mu$), we get:

$$\min_{|\psi^N\rangle, \{M_{\hat{\theta}}\}} \bar{\mathcal{C}} \approx \frac{1}{\bar{N}^2} \min_f \int_0^{+\infty} d\mu f^*(\mu) \left(-\frac{d^2}{d\mu^2} \right) f(\mu), \quad \int_0^{+\infty} |f(\mu)|^2 d\mu = 1, \quad f(0) = 0, \quad \int_0^{+\infty} |f(\mu)|^2 \mu d\mu = 1 \quad (4.26)$$

The solution may be found using the standard Lagrange multiplier method,

$$-\frac{\partial^2}{\partial \mu^2} f(\mu) + f(\mu)(\lambda_1 + \mu \lambda_2) = 0 \Rightarrow f(\mu) \propto \text{Ai} \left(\frac{\lambda_1 + \lambda_2 \mu}{\lambda_2^{2/3}} \right), \quad (4.27)$$

where $\text{Ai}(\cdot)$ is the Airy function of the first kind. After taking into account the conditions, we obtain:

$$f(\mu) = \frac{1}{\text{Ai}'(A_0)} \sqrt{\frac{2|A_0|}{3}} \text{Ai} \left(A_0 + \frac{2|A_0|}{3} \mu \right), \quad (4.28)$$

where $A_0 \approx -2.34$ is the first zero of the Airy function and $\text{Ai}'(\cdot)$ is its first derivative. The

corresponding minimal obtainable cost reads:

$$\min_{|\psi^N\rangle, \{M_{\hat{\theta}}\}} \bar{\mathcal{C}} \approx \frac{1}{N^2} \frac{4|A_0|^3}{27} = \frac{1}{N^2} \underbrace{\frac{4|A_0|^3}{27}}_{c \approx 1.89}. \quad (4.29)$$

One comment should be added here. In the above proof, we restrict ourselves to pure states. Therefore, for completeness, we argue that this cannot be overcome by any mixed input state. For simplicity consider two-rank mixed state $\rho = p_1|\psi_1\rangle\langle\psi_1| + p_2|\psi_2\rangle\langle\psi_2|$, where $\text{Tr}(|\psi_1\rangle\langle\psi_1|H) = \bar{N}_1$, $\text{Tr}(|\psi_2\rangle\langle\psi_2|H) = \bar{N}_2$, with $p_1\bar{N}_1 + p_2\bar{N}_2 = \bar{N}$. The minimal cost obtainable for the mixed input state is bigger or equal to the weighted sum of costs optimal for its pure components, therefore:

$$\bar{\mathcal{C}}[\rho] \geq p_1\bar{\mathcal{C}}[|\psi_1\rangle\langle\psi_1|] + p_2\bar{\mathcal{C}}[|\psi_2\rangle\langle\psi_2|] \geq p_1\frac{c}{\bar{N}_1^2} + p_2\frac{c}{\bar{N}_2^2} \geq \frac{c}{(p_1\bar{N}_1 + p_2\bar{N}_2)^2} = \frac{c}{\bar{N}^2}, \quad (4.30)$$

where by $\bar{\mathcal{C}}[\rho]$ I denoted the minimal variance obtainable for given state ρ . The third inequality comes from the convexity property of the $1/x^2$ function (Jensen inequality). The reasoning may be trivially extended for the ρ of higher rank.

4.5. π -corrected Heisenberg limit

Let us go back to the fixed number of photons N . Looking at the optimal cost [Eq. \(4.21\)](#) and exemplary measurement, which obtains this cost [Eq. \(4.22\)](#), one may have the impression that the resulting value $\approx \frac{\pi^2}{N^2}$ is somehow related to dividing initial interval $[0, 2\pi[$ into $\sim N$ parts corresponding to the measurement outcomes.

However, this intuition fails: in [\[Hayashi \[2011\]\]](#), it was proven that, up to the leading term, the optimal cost achievable in parallel strategy does not depend on the size of the initial interval. More precisely, in the minimax formalism $\min_{\theta \in \Theta} \lim_{N \rightarrow \infty} N^2 \Delta^2 \tilde{\theta} = \pi^2$, independently of the size of Θ (more details about the methods used in this paper will be discussed in [Ch. 6](#)).

However, three issues remain unclear:

- is this results universal (does it works for any unitary generator)?
- how the convergence of the above limit depends on the size of Θ ?
- does the statement also hold in the adaptive scheme?

Below I present the results from [\[Górecki et al. \[2020\]\]](#), which answer all these questions. In the original work, the main thesis has been formulated within Bayesian formalism. Here I present a slightly simplified version formulated within the minimax approach.

Theorem. Given a system with unitary evolution $U_{\theta} = e^{i\theta\Lambda}$ governed by the bounded generator Λ

with extreme eigenvalues λ_- and λ_+ . No further assumptions about the spectrum of Λ are made. Especially the spectrum may be continuous. For more concise notation, I will denote the difference of the extreme eigenvalues by $\lambda = \lambda_+ - \lambda_-$. Then, assuming that from the beginning, the parameter is known to be in the interval, $\theta \in \Theta = [\theta_0 - \delta/2, \theta_0 + \delta/2]$, the optimal minimax cost, obtainable in the most general adaptive scheme with the usage of N quantum gate is bounded by:

$$\widehat{\Delta^2 \tilde{\theta}} \geq \frac{\pi^2}{N^2 \lambda^2} \left(1 - \frac{8 \log(N \lambda \delta)}{N \lambda \delta} \right). \quad (4.31)$$

In particular, $\lim_{N \rightarrow \infty} \widehat{N^2 \Delta^2 \tilde{\theta}} \geq \pi^2 / \lambda^2$, which is saturable within the parallel scheme by the analog of the solution from [Sec. 4.3](#). A more detailed discussion will be performed after the derivation of the bound.

Proof. The proof contains two major parts. Firstly, I will derive the bound for the Bayesian cost for a specific class of prior distribution $p(\theta)$. Next, I will show that, up to small correction, this bound will also be valid for minimax cost for $\theta \in \Theta = [\theta_0 - \delta/2, \theta_0 + \delta/2]$.

First note that from the Bayesian rule, for any measurement, the Bayesian cost may be bounded from below by the minimal variance of a posteriori distribution $p(\theta|x)$

$$\overline{\Delta^2 \tilde{\theta}} = \int d\theta \int dx (\tilde{\theta}_x - \theta)^2 p(x|\theta) p(\theta) = \int d\theta \int dx (\tilde{\theta}_x - \theta)^2 p(\theta|x) p(x) \geq \min_x \int d\theta (\tilde{\theta}_x - \theta)^2 p(\theta|x), \quad (4.32)$$

where we assume optimal $\tilde{\theta}_x$, i.e. the one pointing on mean value of $p(\theta|x)$.

Let name by $|\psi_\theta^N\rangle$ the output state after N acting of gates in the adaptive scheme. According to [Sec. 2.3](#), without loss we may restrict to projective rank-one measurements $p(x|\theta) = |\langle \chi | \psi_\theta^N \rangle|^2$, so:

$$p(\theta|x) = \frac{|\langle \chi | \psi_\theta^N \rangle|^2 p(\theta)}{\int |\langle \chi | \psi_\theta^N \rangle|^2 p(\theta) d\theta} \quad (4.33)$$

Note that scalar function $f(\theta) = \langle \chi | \psi_\theta^N \rangle$ have a finite bandwidth (the support of its Fourier transform $\hat{f}(\mu)$) of the size of $N(\lambda_+ - \lambda_-)$ ¹. It may be clearly seen, when performing the calculation in the eigenbasis of Λ , as dependence on θ will be visible only in terms which are a product of N

¹note that while $\hat{f}(\mu)$ plays a similar role as in the previous section, it is not exactly the same object, as here $\mu \in [N\lambda_-, N\lambda_+]$, not $[\lambda_-, \lambda_+]$

terms $e^{i\mu_k\theta}$ with $\lambda_- \leq \mu_k \leq \lambda_+$:

$$\begin{aligned}
\langle \chi | \psi_\theta^N \rangle &= \\
\langle \chi | V_N \left(\int_{\lambda_-}^{\lambda_+} d\mu_N e^{i\mu_N\theta} |\mu_N\rangle \langle \mu_N| \right) \cdots V_2 \left(\int_{\lambda_-}^{\lambda_+} d\mu_2 e^{i\mu_2\theta} |\mu_2\rangle \langle \mu_2| \right) V_1 \left(\int_{\lambda_-}^{\lambda_+} d\mu_1 e^{i\mu_1\theta} |\mu_1\rangle \langle \mu_1| \right) | \psi_{\text{in}} \rangle \\
&= \int_{\lambda_-}^{\lambda_+} d\mu_N \cdots \int_{\lambda_-}^{\lambda_+} d\mu_1 e^{i(\mu_1 + \cdots + \mu_N)\theta} \langle \chi | V_N | \mu_N \rangle \prod_{k=1}^{N-1} \langle \mu_{k+1} | V_k | \mu_k \rangle \langle \mu_1 | \psi_{\text{in}} \rangle = \\
&\quad \underbrace{\int_{N\lambda_-}^{N\lambda_+} d\mu e^{i\mu\theta} \int d\boldsymbol{\mu}^\perp \langle \chi | V_N | \mu_N \rangle \prod_{k=1}^{N-1} \langle \mu_{k+1} | V_k | \mu_k \rangle \langle \mu_1 | \psi \rangle}_{\theta\text{-independent part}} \quad (4.34)
\end{aligned}$$

where $\mu = \mu_1 + \cdots + \mu_N$, $\int d\boldsymbol{\mu}^\perp$ corresponds to integration over all directions where $\mu = \text{const}$.

Next, to make the problem analytically solvable, let us consider the class of probability distribution $p(\theta)$, which also has finite bandwidth L . It was proven [R. P. Boas and Kac, 1945, Lemma 5.1.], that for any positive function $p_L(\theta)$ of the bandwidth L there exists the function $w_{L/2}(\theta)$ with bandwidth $L/2$ such that $p_L(\theta) = |w_{L/2}(\theta)|^2$. Therefore we have

$$p(\theta|x) = \frac{|f(\theta)w_{L/2}(\theta)|^2}{\int |f(\theta)w_{L/2}(\theta)|^2 d\theta} = |g(\theta)|^2, \quad \text{with} \quad g(\theta) := \frac{f(\theta)w_{L/2}(\theta)}{\sqrt{\int |f(\theta)w_{L/2}(\theta)|^2 d\theta}}, \quad (4.35)$$

where $g(\theta)$, as a (normalized) product of two functions with bandwidths $N\lambda$ and $L/2$, has a bandwidth $N\lambda + L/2$ (as its Fourier transform is equal to the convolution of Fourier transforms of $f(\theta)$ and $w_{L/2}(\theta)$).

Therefore RHS of Eq. (4.32) may be bounded from below by minimization over all $p(\theta|x) = |g(\theta)|^2$ (we may also assume without a loss that its mean value is at $\theta = 0$, as shifting whole distribution does not change the bandwidth):

$$\overline{\Delta^2 \tilde{\theta}} \geq \min_{g: \text{supp } \hat{g} \in [N\lambda_- - L/4, N\lambda_+ + L/4]} \int d\theta |g(\theta)|^2 \theta^2, \quad \int d\theta |g(\theta)|^2 = 1 \quad (4.36)$$

After applying Fourier transform:

$$\overline{\Delta^2 \tilde{\theta}} \geq \min_{\hat{g}} \int_{N\lambda_- - L/4}^{N\lambda_+ + L/4} d\mu \left| \frac{d\hat{g}(\mu)}{d\mu} \right|^2, \quad \int_{-N\lambda_- - L/4}^{N\lambda_+ + L/4} d\mu |g(\mu)|^2 = 1 \quad (4.37)$$

so it comes down to finding the minimum energy eigenstate in an infinite potential well and, therefore

$$\overline{\Delta^2 \tilde{\theta}} \geq \frac{\pi^2}{[N\lambda + L/2]^2}. \quad (4.38)$$

Now, let us go to the second part, where I use the above results to derive the bound for minimax cost with Θ of size δ . It is clear that as p_L has a finite bandwidth, it cannot itself has finite support δ , so it cannot be applied directly from [Eq. \(4.4\)](#). However, one may introduce the finite bandwidth function, for which the probability of finding θ outside of Θ is sufficiently small that it does not affect the final cost in the leading term. An example of such a function is

$$p_{\alpha,L}(\theta) = \mathcal{N}_\alpha L \text{sinc}^4 \left(\pi \alpha \sqrt{(L\theta/4\alpha)^2 - 1} \right), \quad \text{with} \quad \mathcal{N}_\alpha \approx 4\sqrt{2}\pi^4 \alpha^{7/2} e^{-4\pi\alpha} \quad (4.39)$$

which is proportional to the fourth power of the Fourier transform of the Kaiser window [[Kaiser and Schafer \[1980\]](#)]. The width of the main body of such a function is $8\alpha/L$, while the weight of tails decreases exponentially like $e^{-4\pi\alpha}$.

To show it formally, let me introduce compact notation for minimal Bayesian/minimax cost with given set Θ , probability distribution $p(\theta)$ and number of gates N :

$$\text{minimax}(\Theta, N) := \inf_{\rho, V_i, M_i} \sup_{\tilde{\theta} \in \Theta} \int d\tilde{\theta} \text{Tr}(\rho_\theta^N M_{\tilde{\theta}}) (\tilde{\theta} - \theta)^2, \quad (4.40)$$

$$\text{Bayesian}(p(\theta), N) := \inf_{\rho, V_i, M_i} \int p(\theta) d\theta \int d\tilde{\theta} \text{Tr}(\rho_\theta^N M_{\tilde{\theta}}) (\tilde{\theta} - \theta)^2, \quad (4.41)$$

where ρ_θ^N is the output state obtained by acting on the input state ρ the gate N times with unitaries V_i between, as in the scheme [2.3\(c\)](#).

Without loss of generality assume that $\Theta = [-\delta/2, +\delta/2]$ (i.e. $\theta_0 = 0$). Let R_1 be the probability, that θ lays outside of Θ (we consider symmetric $p_L(\theta)$):

$$R_1 = 2 \int_{\delta/2}^{+\infty} d\theta p_L(\theta). \quad (4.42)$$

Next, we define renormalized cut probability distribution:

$$p_L^\delta(\theta) = \begin{cases} \frac{1}{1-R_1} p_L(\theta) & \text{for } \theta \in \Theta \\ 0 & \text{for } \theta \notin \Theta. \end{cases} \quad (4.43)$$

For such distribution from [Eq. \(4.4\)](#):

$$\text{Bayesian}(p_L^\delta, N) \leq \text{minimax}([-\delta/2, +\delta/2], N). \quad (4.44)$$

Let $\rho_\theta^{N,\delta}, M_\theta^\delta$ be the protocol optimal for p_L^δ . Trivially, for uncut probability p_L this protocol is only

suboptimal:

$$\text{Bayesian}(p_L, N) = \inf_{\rho, V_i, M_i} \int_{\mathbb{R}} p_L(\theta) d\theta \int_{\mathbb{R}} d\tilde{\theta} \text{Tr}(\rho_{\theta} M_{\tilde{\theta}}) (\tilde{\theta} - \theta)^2 \leq \int_{\mathbb{R}} p_L(\theta) d\theta \int_{\mathbb{R}} d\tilde{\theta} \text{Tr}(\rho_{\theta}^{N, \delta} M_{\tilde{\theta}}^{\delta}) (\tilde{\theta} - \theta)^2. \quad (4.45)$$

The last integration may be split between two parts $[-\delta/2, +\delta/2]$ and $\mathbb{R} \setminus (-\delta/2, +\delta/2)$

$$\frac{1}{1 - R_1} \underbrace{\int_{-\delta/2}^{\delta/2} p_L^{\delta}(\theta) d\theta \int_{-\delta/2}^{\delta/2} d\tilde{\theta} \text{Tr}(\rho_{\theta}^{N, \delta} M_{\tilde{\theta}}^{\delta}) (\tilde{\theta} - \theta)^2}_{\text{Bayesian}(p_L^{\delta}, N)} + \underbrace{\int_{\mathbb{R} \setminus (-\delta/2, +\delta/2)} p_L(\theta) d\theta \int_{-\delta/2}^{\delta/2} d\tilde{\theta} \text{Tr}(\rho_{\theta}^{N, \delta} M_{\tilde{\theta}}^{\delta}) (\tilde{\theta} - \theta)^2}_{\leq R_2}. \quad (4.46)$$

The first one is simply $\text{Bayesian}(p_L^{\delta}, N)$ rescaled by factor $\frac{1}{1 - R_1}$. The second part may be bounded from above by assuming the most pessimistic scenario – each time the true value of parameter θ is to the left (or right) on $\delta/2$, the indication of the estimator is equal $-\delta/2$ (or $+\delta/2$):

$$R_2 = 2 \int_{\delta/2}^{+\infty} d\theta p_L(\theta) (\theta - \delta/2)^2. \quad (4.47)$$

From [Eq. \(4.45\)](#), [Eq. \(4.46\)](#) we have:

$$\text{Bayesian}(p_L, N) \leq \frac{1}{1 - R_1} \text{Bayesian}(p_L^{\delta}, N) + R_2 \quad (4.48)$$

Replacing $(1 - R_1)$ by one, subtracting R_2 and applying [Eq. \(4.38\)](#) and [Eq. \(4.44\)](#) we get:

$$\frac{\pi^2}{(N\lambda + L/2)^2} - R_2 \leq \text{minimax}([-\delta/2, +\delta/2], N). \quad (4.49)$$

Next, in [App. B](#) I prove that for the function [Eq. \(4.39\)](#) with $L = \frac{8}{\delta} \log(N\lambda\delta)$, $\alpha = \log(N\delta\lambda)$ for $N\lambda\delta > 2$:

$$\frac{\pi^2}{\lambda^2 N^2} \left(1 - \frac{8 \log(N\lambda\delta)}{N\lambda\delta} \right) \leq \frac{\pi^2}{(\lambda N + L/2)^2} - R_2. \quad (4.50)$$

By applying [Eq. \(4.50\)](#) to [Eq. \(4.49\)](#), we get:

$$\frac{\pi^2}{\lambda^2 N^2} \left(1 - \frac{8 \log(N\lambda\delta)}{N\lambda\delta} \right) \leq \text{minimax}([-\delta/2, +\delta/2], N), \quad (4.51)$$

or, in more common notation:

$$\widehat{\Delta^2 \tilde{\theta}} \geq \frac{\pi^2}{\lambda^2 N^2} \left(1 - \frac{8 \log(N\lambda\delta)}{N\lambda\delta} \right), \quad (4.52)$$

what ends the proof. Note, that the function $1 - 8 \log(x)/x$ takes a positive value for $x \geq 26.1$, so

only for $N\lambda\delta \geq 26.1$ above bound is informative. $\square$

In [Górecki et al. [2020]] the analogous bound has been derived in a similar spirit for the Bayesian cost with rectangular flat prior (the second part of the proof required more calculations):

$$\overline{\Delta^2\tilde{\theta}} \geq \frac{\pi^2}{\lambda^2 N^2} \left(1 - \sqrt{\frac{8 \log(N\lambda\delta)}{N\lambda\delta}} \right). \quad (4.53)$$

One may quickly notice that the bound converge slower than the previous one, which stays consistent with Eq. (4.4).

Asymptotic saturability

The exact solution for the estimating unknown phase in the interferometer, namely Eq. (4.20), we can use it to show asymptotical saturability of Eq. (4.52). Therefore, we will show that the bound may be asymptotically saturated within a parallel scheme.

By equating the upper and lower arms of the interferometer with the eigenstates of Λ with extreme eigenvalues, we get the analog of Eq. (4.20):

$$|\psi\rangle = \sqrt{\frac{1}{N+2}} \sum_{m=0}^N \sin\left(\frac{(m+1)\pi}{N+2}\right) |N-m\rangle_{\lambda_+} |m\rangle_{\lambda_-}, \quad (4.54)$$

where $|N-m\rangle_{\lambda_+} |m\rangle_{\lambda_-}$ is a symmetric state with m particles in state $|\lambda_- \rangle$ and $N-m$ particles in state $|\lambda_+ \rangle$. The only potential problem is that this state cannot distinguish the phases, which differ by multiplication of $2\pi/\lambda$. Hence, it allows only for effective estimation inside the region of size $2\pi/\lambda$. However, unlike in the $N00N$ state case, this is not a severe problem, as the size $2\pi/\lambda$ does not shrink with N . Therefore, one may spend at the beginning some sublinear amount of resources (for example $\sqrt{N}$) to discriminate such region (with exponentially decreasing probability of error) without affecting the final cost up to terms of order $1/N^2$.

On the other hand, in a specific case when the spectrum is continuous everywhere between $[\lambda_-, \lambda_+]$, one may use the "continuous" version of the above state, namely

$$|\psi_{\tilde{\theta}}^N\rangle = \int_{N\lambda_-}^{N\lambda_+} e^{i\mu\tilde{\theta}} \sqrt{\frac{2}{\lambda N}} \sin\left(\frac{\mu\pi}{\lambda N}\right) |\mu\rangle d\mu, \quad (4.55)$$

where, by $|\mu\rangle$, we denote an arbitrary combination $|\mu_N\rangle \otimes \dots \otimes |\mu_1\rangle$ such that $\sum \mu_k = \mu$ (i.e. the one satisfying $U_{\tilde{\theta}}^{\otimes N} |\mu\rangle = e^{i\mu\tilde{\theta}} |\mu\rangle$), with covariant measurement:

$$|\tilde{\theta}\rangle = \frac{1}{\sqrt{2\pi}} \int e^{i\mu\tilde{\theta}} |\mu\rangle d\mu, \quad (4.56)$$

to get immediatly $\widehat{\Delta^2\tilde{\theta}} = \pi^2/N^2\lambda^2$ in single shot experiment.

Intuitive interpretation

Once the bound is derived, I would like to provide a simple intuition for the finite bandwidth prior distribution used during the proof. That would be, in fact, an alternative derivation of the bound Eq. (4.38), but only for the parallel strategies [Górecki and Demkowicz-Dobrzański [2022a]].

Let start with the problem, where Λ has continuous spectrum $[\lambda_-, \lambda_+]$ ² and the parameter is entirely unknown $\theta \in \mathbb{R}$ (i.e. $\delta = +\infty$). From the generalization of covariant measurement theorem for non-compact groups [Bogomolov [1982]], we can use the optimality of covariant measurement, and the optimal solution is the one discussed above with cost $\widehat{\Delta^2\tilde{\theta}} = \pi^2/N^2\lambda^2$.

Therefore, assuming that at the beginning one has to disposal $N + N_0$ gates, the minimal achievable minimax cost will be therefore equal:

$$\text{minimax}(\mathbb{R}, N + N_0, \text{parallel}) = \frac{\pi^2}{(N + N_0)^2\lambda^2}. \quad (4.57)$$

On the other hand, we can consider an alternative, sub-optimal strategy: at first use state $|\psi^{N_0}\rangle$ with the measurement Eq. (4.56) to get an approximated value of the parameter $\tilde{\theta}_{\text{est}}$, and after that spend remaining N gates to estimate a difference $\theta - \tilde{\theta}_{\text{est}}$, i.e, use input state

$$|\psi^{N_0+N}\rangle = |\psi^{N_0}\rangle \otimes |\psi^N\rangle, \quad (4.58)$$

with the measurement:

$$|\tilde{\theta}_{\text{est}}\rangle\langle\tilde{\theta}_{\text{est}}| \otimes U_{\tilde{\theta}_{\text{est}}}^{\otimes N} M_{\tilde{\theta}-\tilde{\theta}_{\text{est}}} U_{\tilde{\theta}_{\text{est}}}^{\dagger\otimes N}. \quad (4.59)$$

The resulting minimax cost will be given as:

$$\widehat{\Delta^2\tilde{\theta}} = \sup_{\theta \in \mathbb{R}} \int_{\mathbb{R}} d\tilde{\theta} \int_{\mathbb{R}} d\tilde{\theta}_{\text{est}} |\langle\tilde{\theta}_{\text{est}} - \theta|\psi_0^{N_0}\rangle|^2 \text{Tr} \left(M_{\tilde{\theta}-\tilde{\theta}_{\text{est}}} |\psi_{\theta-\tilde{\theta}_{\text{est}}}^N\rangle\langle\psi_{\theta-\tilde{\theta}_{\text{est}}}^N| \right) (\theta - \tilde{\theta})^2, \quad (4.60)$$

(where we used $\langle\tilde{\theta}_{\text{est}}|\psi_{\theta}^{N_0}\rangle = \langle\tilde{\theta}_{\text{est}} - \theta|\psi_0^{N_0}\rangle$). We see that the cost does not depend on θ , so the supremum at the beginning may be omitted. Introducing $\varphi = \theta - \tilde{\theta}_{\text{est}}$, $\tilde{\varphi} = \tilde{\theta} - \tilde{\theta}_{\text{est}}$ we get

$$\widehat{\Delta^2\tilde{\theta}} = \int_{\mathbb{R}} d\varphi \int_{\mathbb{R}} d\tilde{\varphi} |\langle\varphi|\psi_0^{N_0}\rangle|^2 \text{Tr} (M_{\tilde{\varphi}} U_{\varphi} |\psi_{\varphi}^N\rangle\langle\psi_{\varphi}^N| U_{\varphi}^{\dagger}) (\varphi - \tilde{\varphi})^2, \quad (4.61)$$

so the formula for minimax cost $\widehat{\Delta^2\tilde{\theta}}$ in this peculiar strategy is, in fact, equal to the formula for

²Note that from an estimation point of view, this is an easier problem than one with a discrete spectrum since we have more space in the choice of the input state. Therefore the lower bound derived for the continuous spectrum is also valid for the discrete one.

the Bayesian cost $\overline{\Delta^2 \tilde{\varphi}}$ with the a priori distribution $p(\varphi) = |\langle \varphi | \psi_0^{N_0} \rangle|^2$! Moreover, for the state:

$$|\psi_\theta^{N_0}\rangle = \int_{N_0\lambda_-}^{N_0\lambda_+} e^{i\mu\theta} \hat{w}_{\lambda N_0}(\mu) |\mu\rangle d\mu, \quad (4.62)$$

the probability density of obtaining results $\tilde{\theta}_{\text{est}}$ is:

$$p(\tilde{\theta}_{\text{est}}|\theta) = |\langle \tilde{\theta}_{\text{est}} | \psi \rangle|^2 = |w_{\lambda N_0}(\theta - \tilde{\theta}_{\text{est}})|^2 \quad (4.63)$$

Therefore, one get a posteriori distribution $p_{2\lambda N_0}(\varphi) = |w_{\lambda N_0}(\theta - \tilde{\theta}_{\text{est}})|^2$, with the bandwidth $L = 2\lambda N_0$. That it is – the finite bandwidth probability distribution of the bandwidth L is the one which may be obtained starting from a completely unknown parameter by spending $N_0 = L/(2\lambda)$ gates! As this two-step procedure is only sub-optimal, we immediately get:

$$\text{bayesian}(p_{2\lambda N_0}, N, \text{parallel}) \geq \text{minimax}(\mathbb{R}, N + N_0, \text{parallel}), \quad (4.64)$$

which leads to the same bound formula as [Eq. \(4.52\)](#), but here for the parallel strategies only:

$$\overline{\Delta^2 \tilde{\varphi}} \geq \frac{\pi^2}{[N\lambda + L/2]^2}. \quad (4.65)$$

The reasoning could be performed for the adaptive scheme as well, but then one would need to independently prove [Eq. \(4.57\)](#) for the general adaptive scheme. This makes the first approach more straightforward.

4.6. Average energy constraint

In this section, I will derive the analogous bound for the problem of the unitary estimation with average energy constraint, discussed in [Sec. 2.4](#). The bound is a consequence of discussions and calculations included in supplementary materials of [\[Górecki and Demkowicz-Dobrzański \[2022a,b\]\]](#), but it was not written explicitly anywhere before.

First, note that we cannot simply repeat the main reasoning from [Sec. 4.5](#). Indeed, having constraints only on average energy, one can easily find an example where RHS in the inequality [Eq. \(4.32\)](#) goes to 0 for the fixed value of average energy. However, we may make use of the comment made at the end (about an intuitive interpretation) to derive the bound for parallel strategy (with ancilla). For more compact notation, let us assume that the ground state energy is zero $E_g = 0$. The problem is defined as follows:

$$e^{i\theta H}, \quad \langle \psi | H | \psi \rangle \leq E, \quad \theta \in [\theta_0 - \delta/2, \theta_0 + \delta/2] \quad (4.66)$$

and we want to minimize the MSE of the estimator of θ . Without loss, we assume that the

spectrum of H is continuous $[0, +\infty[$ (allowing for continuous spectrum may only decrease the minimal obtainable cost, as there is more freedom to choose input state). Then we can use the reasoning mentioned at the end of the previous section. Consider first the problem of estimating a completely unknown parameter $\theta \in \mathbb{R}$, using the state with the average energy $E + E_0$. Then, from the reasoning performed in [Sec. 4.4](#),

$$\widehat{\Delta^2 \tilde{\theta}} \geq \frac{4|A_0|^3}{27} \frac{1}{(E + E_0)^2}. \quad (4.67)$$

Next, we consider a suboptimal two-step strategy. Consider the input state of the form:

$$|\psi^{E+E_0}\rangle = |\psi^{\leq E_0}\rangle \otimes |\psi^E\rangle, \quad (4.68)$$

where $|\psi^{\leq E_0}\rangle$ is defined as:

$$|\psi_{\tilde{\theta}}^{\leq E_0}\rangle = \int_0^{E_0} e^{i\mu\tilde{\theta}} \hat{w}_{E_0}(\mu) |\mu\rangle d\mu, \quad (4.69)$$

so for sure $\langle \psi | H | \psi \rangle \leq E_0$. Then we perform a covariant measurement on $e^{i\theta H} |\psi^{\leq E_0}\rangle$. The bandwidth of the resulting probability distribution:

$$p(\tilde{\theta}|\theta) = |\langle \tilde{\theta} | \psi_{\tilde{\theta}}^{\leq E_0} \rangle|^2 = |w_{E_0}(\theta - \tilde{\theta})|^2 \quad (4.70)$$

is therefore $L = 2E_0$. Using the argument that such a two-step strategy is only suboptimal, in an analog to the previous case, we get:

$$\text{bayesian}(p_{2E_0}, E, \text{parallel}) \geq \text{minimax}(\mathbb{R}, E + E_0, \text{parallel}) \quad (4.71)$$

$$\overline{\Delta^2 \tilde{\theta}} \geq \frac{4|A_0|^3}{27} \frac{1}{(E + L/2)^2}. \quad (4.72)$$

Then, choosing as $p_{2E_0}(\theta)$ the function [Eq. \(4.39\)](#), and repeating reasoning performed in [4.5](#), we obtain:

$$\widehat{\Delta^2 \tilde{\theta}} \geq \frac{4|A_0|^3}{27} \frac{1}{E^2} \left(1 - \frac{8 \log(E\delta)}{E\delta} \right). \quad (4.73)$$

As mentioned before, the function $1 - 8 \log(x)/x$ takes a positive value for $x \geq 26.1$, so only for $E\delta \geq 26.1$ above bound is informative. For $E\delta \rightarrow \infty$ it converge to $\frac{4|A_0|^3}{27} \frac{1}{E^2}$ and, in contrast to existing bounds mentioned in [Sec. 2.4](#), it is asymptotically saturable.

CHAPTER 5

MULTIPARAMETER METROLOGY

5.1. Foundations

In the previous chapters, I discussed the situation where the quantum channel depends on a single (scalar) unknown parameter we want to measure. On the other hand, many practical metrology problems are related to the estimation of multiple parameters simultaneously, such as vector field sensing (e.g., magnetic field) [Baumgratz and Datta [2016]], imaging [Tsang et al. [2016]], multiple-arm interferometry [Humphreys et al. [2013b], Gessner et al. [2018]] or waveform estimation [Tsang et al. [2011], Berry et al. [2013]], which makes multi-parameter metrology an object of intensive interest in recent years [Kolenderski and Demkowicz-Dobrzanski [2008], Matsumoto [2002], Genoni et al. [2013], Ragy et al. [2016], Yuan [2016], Kura and Ueda [2018], Liu and Yuan [2017], Nichols et al. [2018], Ge et al. [2018]]. Let me formulate the problem more concretely.

Let $\mathcal{E}_{\boldsymbol{\theta}}(\cdot)$ be the quantum channel depending on the vector of the parameters $\boldsymbol{\theta} = [\theta_1, \dots, \theta_p]^T$, and so does the results of the measurement performed on the output state $p(x|\boldsymbol{\theta}) = \text{Tr}(M_x \rho_{\boldsymbol{\theta}})$.

The aim is, therefore, to estimate the values $\boldsymbol{\theta}$, by probing x . We introduce the cost function $\mathcal{C}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})$ which is to be minimized:

$$\mathcal{C}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) = \sum_x \mathcal{C}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}(x)) p(x|\boldsymbol{\theta}), \quad (5.1)$$

which, for a typical problem, for the estimator indication close to the true value may be well approximated by the quadratic function:

$$\mathcal{C}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \approx (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^T C (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}), \quad (5.2)$$

where C is the Hessian of the cost function $C_{ij} = \partial_i \partial_j \mathcal{C}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}})|_{\tilde{\boldsymbol{\theta}}=\boldsymbol{\theta}}$ (which, in general, may depend on the exact value of $\boldsymbol{\theta}$). Here ∂_i is the shortcut for $\frac{\partial}{\partial \theta_i}$. It is convenient to write down the average cost as:

$$\text{tr}(C\Sigma), \quad \text{where} \quad \Sigma := \sum_x p(x|\boldsymbol{\theta}) (\tilde{\boldsymbol{\theta}}(x) - \boldsymbol{\theta})(\tilde{\boldsymbol{\theta}}(x) - \boldsymbol{\theta})^T, \quad (5.3)$$

where $\text{tr}(\cdot)$ denote trace over p -dimensional space of parameters (I use the small letter $\text{tr}(\cdot)$ to distinguish from the trace over physical Hilbert space $\text{Tr}(\cdot)$). If the estimator is unbiased at $\boldsymbol{\theta}$, i.e. $\sum_x p(x|\boldsymbol{\theta}) \tilde{\boldsymbol{\theta}}(x) = \boldsymbol{\theta}$, Σ is the covariance matrix of the $\tilde{\boldsymbol{\theta}}$. Especially if the cost matrix is identity $C = \mathbb{1}$, the cost is simply the sum of variances of the estimators $\sum_{i=1}^p \Delta^2 \tilde{\theta}_i$.

5.2. Fisher information matrix and multiparameter Cramér-Rao bound

Let me start with the scenario where the experiment is repeated many times so that we may use Fisher information formalism.

Given probability distribution $p(x|\boldsymbol{\theta})$. By locally unbiased estimator of the component θ_i , we understand the one which does not only properly changes with arising θ_i , but is also insensitive for changes of the remaining parameters $\theta_{i \neq j}$. For any estimator locally unbiased at $\boldsymbol{\theta}$:

$$\sum_x p(x|\boldsymbol{\theta}) \tilde{\theta}_j(x) = \theta_j, \quad \sum_x \frac{\partial p(x|\boldsymbol{\theta})}{\partial \theta_i} \tilde{\theta}_j(x) = \delta_{ij} \quad (5.4)$$

the covariance matrix is bounded from below by multiparameter Cramér-Rao bound

$$\Sigma \geq \mathbf{F}^{-1}, \quad \mathbf{F}_{ij} = \sum_x \frac{1}{p(x|\boldsymbol{\theta})} \frac{\partial p(x|\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial p(x|\boldsymbol{\theta})}{\partial \theta_j}, \quad (5.5)$$

which may be derived by using the Cauchy-Schwarz inequality (by matrix inequality, we understand that their difference is positive semidefinite $\Sigma - \mathbf{F}^{-1} \geq 0$). Note that diagonal matrix element $\mathbf{F}_{ii}$ is single-parameter Fisher information for θ_i . However, from Eq. (5.5), $\Delta^2 \tilde{\theta}_i \geq [\mathbf{F}^{-1}]_{ii}$ (where in general $[\mathbf{F}^{-1}]_{ii} \geq (\mathbf{F}_{ii})^{-1}$), as stronger locally unbiasedness condition has been imposed here Eq. (5.4). Similarly, as in the single-parameter case, the Fisher information matrix scales linearly with the number of repetitions k , and in the limit of $k \rightarrow \infty$ the corresponding classical Cramér-Rao inequality is saturated by maximum likelihood estimator [Nagaoka and Hayashi, 2005, CHAPTER 7].

In quantum mechanics, the probability distribution is given by the Born rule $p(x|\boldsymbol{\theta}) = \text{Tr}(\rho_{\boldsymbol{\theta}} M_x)$. Here, I present the derivation of quantum Cramér-Rao bound, based on direct optimization over locally unbiased observable (based on observation from [Ragy et al. [2016]]). The advantage of such derivation is that it clearly shows the origin of the measurement incompatibility problem. Moreover, it allows quick improvement to the stronger bounds [Holevo [1982], Demkowicz-Dobrzański et al. [2020]].

For compact notation, I will denote the vectors of the operators by properly bounded letters. For fixed $\boldsymbol{\theta}$ let me defined:

$$\mathbf{X} = \sum_x (\tilde{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}) M_x, \quad (5.6)$$

i.e., $\mathbf{X} = [X_1, \dots, X_p]^T$, where $X_i = \sum_x (\tilde{\theta}_i(x) - \theta_i) M_x$. We construct a positive semidefinite operator:

$$0 \leq \sum_x [(\tilde{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}) \mathbb{1} - \mathbf{X}] M_x [(\tilde{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}) \mathbb{1} - \mathbf{X}]^T = \sum_x M_x [(\tilde{\boldsymbol{\theta}}(x) - \boldsymbol{\theta})][(\tilde{\boldsymbol{\theta}}(x) - \boldsymbol{\theta})]^T - \mathbf{X} \mathbf{X}^T. \quad (5.7)$$

The first inequality is tight if $\{M_x\}$ is a protective measurement. After applying the trace with density matrix $\text{Tr}(\rho_{\boldsymbol{\theta}} \cdot)$ we obtain:

$$\Sigma \geq \text{Tr}(\rho_{\boldsymbol{\theta}} \mathbf{X} \mathbf{X}^T) =: Z_{\boldsymbol{\theta}}[\mathbf{X}] \quad (5.8)$$

Therefore, for any cost function C , the average cost is bounded by:

$$\text{tr}(C\Sigma) \geq \text{tr}(CZ_\theta[\mathbf{X}]). \quad (5.9)$$

RHS may be directly minimized over all vectors of hermitian matrices $\mathbf{X}$ satisfying locally unbiased conditions $\text{Tr}(\partial_{\theta_i} X_j) = \delta_{ij}$ (we ignore for a while the fact that all X_i should come from the same measurement $\{M_x\}$). Minimization leads to $\mathbf{X} = \mathbf{F}_Q^{-1} \mathbf{L}$, where $\mathbf{L} = [L_1, \dots, L_p]^T$ is a vector of symmetric logarithmic derivatives:

$$\partial_i \rho_\theta = \frac{1}{2} (L_i \rho_\theta + \rho_\theta L_i) \quad (5.10)$$

and $\mathbf{F}_Q$ is the quantum Fisher information matrix given as:

$$\mathbf{F}_Q = \text{ReTr}(\rho_\theta \mathbf{L} \mathbf{L}^T). \quad (5.11)$$

Therefore, the cost is bounded from below:

$$\text{tr}(C\Sigma) \geq \text{tr}(C\mathbf{F}_Q^{-1}). \quad (5.12)$$

What may be surprising, the $\mathbf{X}$ minimizing [Eq. \(5.9\)](#) does not depend on peculiar matrix cost. As the inequality holds for any C , it may be written as matrix inequality:

$$\Sigma \geq \mathbf{F}_Q^{-1}. \quad (5.13)$$

Let us discuss the saturability of the above inequality now. Note that it comes down to the question if there exists a single protective measurement $\{M_x\}$, for which all $X_i = [\mathbf{F}_Q^{-1} \mathbf{L}]_i$ may be written as [Eq. \(5.6\)](#). If so, the classical Fisher information for this measurement is equal $\mathbf{F}_Q$, and in the many repetitions scenario, one may use an argument about the asymptotic saturability of classical CR.

Such measurement always exists if all SLDs mutually commute

$$\forall_{ij} [L_i, L_j] = 0, \quad (5.14)$$

then the such measurement is simply a projection onto a common eigenbasis of all L_i . That will be the case for most of the examples discussed in this thesis. However, for completeness, let me also discuss the opposite situation.

If $\exists_{ij} [L_i, L_j] \neq 0$, the situation becomes to be more complicated; however, some improvement of [Eq. \(5.12\)](#) may be performed. Note that matrix $Z_\theta[\mathbf{X}]$ in [Eq. \(5.8\)](#) in general may be complex, while covariance matrix Σ is always purely real. As $Z_\theta[\mathbf{X}]$ is hermitian, its imaginary part is antisymmetric. Therefore, after tracing both sides of [Eq. \(5.8\)](#) with a real positive cost matrix, we

completely ignore the information hidden in $\text{Im}Z_\theta[\mathbf{X}]$, as $\text{Tr}(C\text{Im}Z_\theta[\mathbf{X}]) = 0$. We can avoid this problem by, instead of minimizing RHS of Eq. (5.12) directly, performing double minimization:

$$\text{tr}(C\Sigma) \geq \min_{V, \mathbf{X}} \text{tr}(CV) : V \geq \text{tr}(\rho_\theta \mathbf{X}\mathbf{X}^T). \quad (5.15)$$

Moreover, the optimization over V for fixed $\mathbf{X}$ may be performed analytically, giving so called Holevo Cramér-Rao (HCR) bound:

$$\text{tr}(C\Sigma) \geq \min_{\mathbf{X}} \text{tr}(CZ[\mathbf{X}]) + \|\sqrt{C}\text{Im}Z[\mathbf{X}]\sqrt{C}\|_1. \quad (5.16)$$

The last inequality has been proven to be asymptotically saturable for many copies of the system if collective measurements on all copies are allowed, i.e., $p(x|\boldsymbol{\theta}) = \text{Tr}(M_x \rho_{\boldsymbol{\theta}}^{\otimes k})$ (see [Demkowicz-Dobrzański et al. [2020]] for both general proof and simple examples).

From that, we can see that if collective measurements are allowed, the condition for saturability Eq. (5.14) may be weakened – indeed, if SLDs do not commute in general, but commute on state, i.e.

$$\text{Tr}(\rho_\theta [L_i, L_j]) = 0, \quad (5.17)$$

then for $\mathbf{X} = \mathbf{F}_Q^{-1}\mathbf{L}$, the operator $\text{Im}Z_\theta[\mathbf{X}] = 0$, so Holevo-Cramer Rao coincidence with standard CR, which means that also the latter one is saturable. Therefore, condition Eq. (5.17) is a necessary and sufficient condition for asymptotically saturating standard multiparameter CR bound with collective measurement. Moreover, in [Matsumoto [2002]], it has been proven that collective measurements are necessary only for the general case of mixed states. In contrast, for the pure state, HCR may be saturated with standard local measurements on single copies of the system.

At last, when the cost matrix C is rank-one, the problem may be understood that there is only one parameter to be measured (or some concrete linear combination of the parameters), but the experimentalist needs to make sure that fluctuations of others do not affect the results (there are called nuisance parameters in this context [Suzuki [2020], Suzuki et al. [2020]]). In such a case, the standard CR is always equivalent to HCR [Demkowicz-Dobrzański et al., 2020, Sec. 2.7], so it is saturable with collective measurement (however, it is not clear if collective measurements are truly necessary).

Adaptive scheme maximizing QFI

In the single parameter case, it was shown for the unitary estimation that the optimal adaptive scheme is sequential acting of the gates n times, with no necessity of unitary control between Sec. 3.5 [Giovannetti et al. [2006]]. Moreover, it offers no advantage over the optimal parallel scheme.

However, this no longer applies to multiparameter estimation, where the optimal unitary control is the one that reverses evolution after each step. Moreover, in some cases, it may overcome the

optimal parallel scheme. The reason for the difference is that in the multiparameter case, the gates with different values of parameter do not necessarily commute $[U_{\theta_0}, U_{\theta}] \neq 0$, which makes the reversing evolution between each gates something diametrically different from reversing it at the end of the whole procedure. The exemplary problem illustrating this effect will be discussed in [Sec. 6.4](#).

This observation was made in [\[Yuan \[2016\]\]](#) with a formal proof of optimality only for a specific case of estimation of the magnetic field sensed by spin-1/2 particle; the proof was based on the connection of QFI with Bures metric. Below I present an alternative and fully general proof (which, best to my knowledge, does not exist in literature).

It should be emphasized that the following reasoning is based only on comparing QFI matrices, so it does not consider the potential incompatibility of the measurements. The question of whether an analogous protocol optimizes the HCR inequality has not been investigated.

Theorem. In the multiparameter unitary estimation problem with evolution U_{θ} at point $\theta = \theta_0$, for any sequential-adaptive strategy $|\psi_{\theta}\rangle = V_n(U_{\theta} \otimes \mathbb{1}) \dots V_1(U_{\theta} \otimes \mathbb{1})|\psi\rangle$, there exists an alternative strategy with $\forall_i V_i = U_{\theta_0}^{\dagger} \otimes \mathbb{1}$ and input state $|\psi'\rangle \in \mathcal{H}_S \otimes \mathcal{H}_A$, where an ancillary system of the same size as the original one, $\dim(\mathcal{H}_S) = \dim(\mathcal{H}_A)$, for which the QFI matrix is bigger or equal to the QFI for the first strategy:

$$\exists_{|\psi'\rangle \in \mathcal{H}_S \otimes \mathcal{H}_A} \quad \mathbf{F}_Q \left[(U_{\theta_0}^{\dagger} U_{\theta} \otimes \mathbb{1})^{\otimes n} |\psi'\rangle \right] \geq \mathbf{F}_Q \left[V_n(U_{\theta} \otimes \mathbb{1}) \dots V_1(U_{\theta} \otimes \mathbb{1}) |\psi\rangle \right]. \quad (5.18)$$

Proof. For any adaptive strategy, the output state after n steps is given as:

$$\rho_{\theta} = V_n(U_{\theta} \otimes \mathbb{1}) V_{n-1}(U_{\theta} \otimes \mathbb{1}) \dots V_1(U_{\theta} \otimes \mathbb{1}) \rho(U_{\theta} \otimes \mathbb{1})^{\dagger} V_1^{\dagger} \dots (U_{\theta} \otimes \mathbb{1})^{\dagger} V_n^{\dagger} \quad (5.19)$$

and its QFI matrix is:

$$\mathbf{F}_Q = \text{Tr}(\rho_{\theta} \mathbf{L} \mathbf{L}^T), \quad \partial_j \rho_{\theta} \big|_{\theta_0} = \frac{1}{2} (\rho_{\theta_0} L_j + L_j \rho_{\theta_0}). \quad (5.20)$$

I introduce $W_i = V_i(U_{\theta_0} \otimes \mathbb{1})$ (so $V_i = W_i(U_{\theta_0} \otimes \mathbb{1})^{\dagger}$). Then:

$$\rho_{\theta} \big|_{\theta_0} = W_n W_{n-1} \dots W_1 \rho W_1^{\dagger} \dots W_n^{\dagger}. \quad (5.21)$$

while its derivatives may be written in the form of the sum of terms when the derivative "hit" i^{th} acting of the channel:

$$\partial_j \rho_{\boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} = \sum_{i=1}^n \partial_j W_n \dots W_i U_{\boldsymbol{\theta}_0}^\dagger U_{\boldsymbol{\theta}} \rho^{(i)} U_{\boldsymbol{\theta}}^\dagger U_{\boldsymbol{\theta}_0} W_i^\dagger \dots W_n^\dagger \Big|_{\boldsymbol{\theta}_0}, \quad \text{with} \quad \rho^{(i)} := W_{i-1} W_{i-2} \dots W_1 \rho W_1^\dagger \dots W_{i-1}^\dagger. \quad (5.22)$$

We define $L_j^{(i)}$ as SLD corresponding to i^{th} "hitting":

$$\partial_j W_n \dots W_i U_{\boldsymbol{\theta}_0}^\dagger U_{\boldsymbol{\theta}} \rho^{(i)} U_{\boldsymbol{\theta}}^\dagger U_{\boldsymbol{\theta}_0} W_i^\dagger \dots W_n^\dagger = \frac{1}{2} (L_j^{(i)} \rho_{\boldsymbol{\theta}_0} + \rho_{\boldsymbol{\theta}_0} L_j^{(i)}) \quad (5.23)$$

and then full SLD is simply sum of $L_j^{(i)}$:

$$L_j = \sum_{i=1}^n L_j^{(i)}. \quad (5.24)$$

Now I will show that at least the same QFI matrix may be obtained by preparing the state being a mixture of $\rho^{(i)}$ entangled with additional ancilla of the size $\dim(\mathcal{H}_{A'}) = n$:

$$\rho' = \frac{1}{n} \sum_{i=1}^n \rho^{(i)} \otimes |i\rangle\langle i|, \quad (5.25)$$

on which we act:

$$(U_{\boldsymbol{\theta}_0}^\dagger U_{\boldsymbol{\theta}})^n \otimes \mathbb{1} \quad (5.26)$$

so

$$\rho'_{\boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n [(U_{\boldsymbol{\theta}_0}^\dagger U_{\boldsymbol{\theta}})^n \otimes \mathbb{1}] \rho^{(i)} [(U_{\boldsymbol{\theta}_0}^\dagger U_{\boldsymbol{\theta}})^n \otimes \mathbb{1}]^\dagger \otimes |i\rangle\langle i|. \quad (5.27)$$

To show that, I introduce additional matrix $\rho''_{\boldsymbol{\theta}}$, constructed by acting on $\rho'_{\boldsymbol{\theta}}$ with unitary control

$$\sum_{i=1}^n W_n \dots W_i \otimes |i\rangle\langle i|. \quad (5.28)$$

As it is $\boldsymbol{\theta}$ -independent unitary, it does not change the QFI (so $\mathbf{F}'_Q = \mathbf{F}''_Q$), but the SLDs of $\rho_{\boldsymbol{\theta}}$ and $\rho''_{\boldsymbol{\theta}}$ will be much simpler to compare. We have:

$$\begin{aligned} \partial_j \rho''_{\boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} &= \frac{1}{n} \sum_{i=1}^n \partial_j W_n \dots W_i \left[(U_{\boldsymbol{\theta}_0}^\dagger U_{\boldsymbol{\theta}})^n \otimes \mathbb{1} \right] \rho^{(i)} \left[(U_{\boldsymbol{\theta}_0}^\dagger U_{\boldsymbol{\theta}})^n \otimes \mathbb{1} \right]^\dagger W_i^\dagger \dots W_n^\dagger \otimes |i\rangle\langle i| \Big|_{\boldsymbol{\theta}_0} = \\ &= \frac{1}{\mathcal{N}} \sum_{i=1}^n \mathcal{N} \partial_j W_n \dots W_i \left[(U_{\boldsymbol{\theta}_0}^\dagger U_{\boldsymbol{\theta}}) \otimes \mathbb{1} \right] \rho^{(i)} \left[(U_{\boldsymbol{\theta}_0}^\dagger U_{\boldsymbol{\theta}}) \otimes \mathbb{1} \right]^\dagger W_i^\dagger \dots W_n^\dagger \otimes |i\rangle\langle i| \Big|_{\boldsymbol{\theta}_0}, \end{aligned} \quad (5.29)$$

so $\rho''_{\boldsymbol{\theta}}$ satisfies:

$$\partial_j \rho''_{\boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=0} = \frac{1}{2} (L_j'' \rho_0'' + \rho_0'' L_j'') \quad (5.30)$$

with

$$\mathbf{L}'' = \sum_{i=1}^n n \mathbf{L}^{(i)} \otimes |i\rangle\langle i|. \quad (5.31)$$

Therefore:

$$\mathbf{F}_Q'' = \text{Tr}(\rho''_{\boldsymbol{\theta}_0} \mathbf{L}'' \mathbf{L}''^T) = n \sum_{i=1}^n \text{Tr}(\rho_{\boldsymbol{\theta}_0} \mathbf{L}^{(i)} \mathbf{L}^{(i)T}) \quad (5.32)$$

In general, for any hermitian operators:

$$n \left(\sum_{i=1}^n \mathbf{L}^{(i)} \mathbf{L}^{(i)T} \right) - \left(\sum_{i=1}^n \mathbf{L}^{(i)} \right) \left(\sum_{i=1}^n \mathbf{L}^{(i)} \right)^T = \frac{1}{2} \sum_{i,j=1}^n \left(\mathbf{L}^{(i)} - \mathbf{L}^{(j)} \right) \left(\mathbf{L}^{(i)} - \mathbf{L}^{(j)} \right)^T \geq 0, \quad (5.33)$$

which, after applying $\text{Tr}(\rho_{\boldsymbol{\theta}_0} \cdot)$ gives:

$$\mathbf{F}_Q'' - \mathbf{F}_Q \geq 0 \Leftrightarrow \mathbf{F}_Q' - \mathbf{F}_Q \geq 0. \quad (5.34)$$

To perform the above theoretical construction, one needs an extremely large ancilla. Let me now argue that at least the same QFI matrix may be obtained with the ancilla of the same size as the physical system $\mathcal{H}_S$. Consider a purification of ρ' , namely $|\psi'\rangle \in \mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_{A'} \otimes \mathcal{H}_p$ (where $\dim(\mathcal{H}_p) = \dim(\mathcal{H}_S \otimes \mathcal{H}_A \otimes \mathcal{H}_{A'})$). It is clear that the QFI matrix of $|\psi'\rangle = (U_{\boldsymbol{\theta}_0} U_{\boldsymbol{\theta}})^n \otimes \mathbb{1} |\psi'\rangle$ is larger or equal to QFI matrix of $\rho_{\boldsymbol{\theta}}$, as the second one may be obtained from the first by tracing over $\mathcal{H}_p$. On the other hand, from Schmidt decomposition:

$$|\psi'\rangle = \sum_{k=1}^{\dim(\mathcal{H}_S)} \eta_k \underbrace{|\psi'_{k,S}\rangle}_{\in \mathcal{H}_S} \otimes \underbrace{|\psi'_{k,A}\rangle}_{\in \mathcal{H}_A \otimes \mathcal{H}_{A'} \otimes \mathcal{H}_p}, \quad (5.35)$$

so it is enough to consider $|\psi'\rangle \in \mathcal{H}_S \otimes \text{span}\{|\psi'_{k,A}\rangle\}_{k=1}^{\dim(\mathcal{H}_S)}$, what was to be proven. $\square$

It also shows that if, alternatively, $\forall [U_{\boldsymbol{\theta}_0}^\dagger, U_{\boldsymbol{\theta}}^\dagger] = 0$, no unitary feedback is needed, and the optimal controls may be chosen as $\mathbb{1}$. That also implies that for such cases, the general sequential adaptive scheme offers no advantage over the optimal parallel one.

Indeed, let $\{|i\rangle\}$ be the common eigenbasis of all generators Λ_k , such that $\Lambda_k |i\rangle = \lambda_i^{(k)} |i\rangle$. Then n sequential acting on the state:

$$U_{\boldsymbol{\theta}}^n \sum_{i=1}^d c_i |i\rangle = \sum_{i=1}^d e^{in \sum_k \theta_k \lambda_i^{(k)}} c_i |i\rangle \quad (5.36)$$

is equivalent to:

$$U_{\boldsymbol{\theta}}^{\otimes n} \sum_{i=1}^d c_i |i\rangle^{\otimes n} = \sum_{i=1}^d e^{in \sum_k \theta_k \lambda_i^{(k)}} c_i |i\rangle^{\otimes n}. \quad (5.37)$$

General theorem for noisy multiparameter estimation within QFI formalism

The “Hamiltonian-not-in-Kraus-span” (HNKS) condition Eq. (3.60) may be easily generalized for a multiparameter case. It was first done for the problem of continuous evolution given by master Lindblad equation [Górecki et al. [2020]], but methodology may be easily adapted to finite-step gates:

$$\mathcal{E}_{\boldsymbol{\theta}}(\rho) = \sum_{k=0}^r K_k(\boldsymbol{\theta}) \rho K_k^\dagger(\boldsymbol{\theta}). \quad (5.38)$$

For each parameter we define corresponding Hamiltonian $\tilde{\Lambda}_i$ and extended Kraus-span $\mathcal{S}_i$ (where compered to $\mathcal{S}$, it additionally contain all remaining Hamiltonian $\tilde{\Lambda}_{j \neq i}$):

$$\tilde{\Lambda}_i = -i \sum_k K_k^\dagger \partial_i K_k, \quad \mathcal{S}_i = \text{span}_{\mathbb{H}}\{\tilde{\Lambda}_{j \neq i}, K_j^\dagger K_k \text{ for all } j, k\}. \quad (5.39)$$

Then Heisenberg scaling in estimation off all p parameters is possible iff $\forall_i \tilde{\Lambda}_i \notin \mathcal{S}_i$. When this condition is satisfied, each parameter may be measured separately by constructing quantum error correction protocol as in single-parameter case [Zhou and Jiang [2021]], where remaining Hamiltonians $\tilde{\Lambda}_{j \neq i}$ are treated as the noise we want to protect again. Otherwise, if the condition is violated, it means that there exists some linear combinations of parameter θ_i , which cannot be estimated with HS precision.

While if and only if conditions for obtaining HS are known, the general asymptotically saturable bound is not known exactly. The methodology used to derive the saturable bounds in single-parameter estimation, presented in Sec. 3.7, cannot be trivially extended for multiparameter case, mainly due to the difficulties connected with inverting QFI matrix and probe incompatibility. See [Albarelli and Demkowicz-Dobrzański [2022]] for recent progress in this direction.

5.3. Bayesian and minimax approaches

Similarly, as in the single parameter case, we may formulate formulas for Bayesian cost:

$$\bar{\mathcal{C}} = \int p(\boldsymbol{\theta}) d\boldsymbol{\theta} \int d\tilde{\boldsymbol{\theta}} \text{Tr}(\rho_{\boldsymbol{\theta}} M_{\tilde{\boldsymbol{\theta}}}) \mathcal{C}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}), \quad (5.40)$$

and and minimax cost:

$$\hat{\mathcal{C}} = \sup_{\boldsymbol{\theta} \in \Theta} \int d\tilde{\boldsymbol{\theta}} \text{Tr}(\rho_{\boldsymbol{\theta}} M_{\tilde{\boldsymbol{\theta}}}) \mathcal{C}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}). \quad (5.41)$$

Again, if $\text{supp } p(\boldsymbol{\theta}) \in \Theta$, the minimax cost bigger than Bayesian one:

$$\hat{\mathcal{C}} \geq \bar{\mathcal{C}}. \quad (5.42)$$

In the limit of many copies k , if collective measurement is allowed, in general, it coincides with Holevo Cramer-Rao (so also with standard Cramer-Rao if $\text{Tr}(\rho_{\boldsymbol{\theta}}[L_i, L_j]) = 0$), however, the formal reasoning is much more complicated and proven fully formally only for specific cases [Gill and Levit [1995], Gill [2008], Guță and Kahn [2006], Guță et al. [2008]].

Covariant problem

In this section, we recall the theorem introduced and proven in [Chiribella et al. [2008b]], saying that for the covariant problem of the group element estimation, the optimal strategy may be found as the parallel covariant one, i.e., there is no advantage in considering a more general sequential adaptive scheme. The proof from [Chiribella et al. [2008b]] is based on the formalism of quantum combs introduced in [Chiribella et al. [2008a]]. For completeness, I recall here the proof, restricting only to what is strictly necessary for the proof.

Consider a problem where the family of channels depending on the parameter is a unitary representation of a compact Lie group G , such that $\mathcal{E}_g(\rho) = U_g \rho U_g^\dagger$ and the aim is to estimate the group element $g \in G$. Then we say, that the estimation problem is covariant if it satisfies the conditions stated in Sec. 4.2, namely:

- the cost function is invariant under the action of the group $\mathcal{C}(g_1, g_2) = \mathcal{C}(hg_1, hg_2)$.
- (required in Bayesian formalism) a priori distribution is invariant under the action of the group $p(g)dg = p(hg)d(hg)$. For simplicity of the notation, I assume that dg is the normalized Haar measure, $\int_G dg = 1$.

We can now formulate the stronger version of Eq. (4.11).

Theorem. For the covariant group estimation problem, as the optimal strategy using N gates one may always choose the parallel one with covariant measurement, i.e. the one of the form:

$$M_g = (U_g^{\dagger \otimes N} \otimes \mathbb{1}) M_e (U_g^{\otimes N} \otimes \mathbb{1}), \quad \int d\tilde{g} (U_{\tilde{g}}^{\dagger \otimes N} \otimes \mathbb{1}) M_e (U_{\tilde{g}}^{\otimes N} \otimes \mathbb{1}) = \mathbb{1}. \quad (5.43)$$

It cannot be overcome even by the most general sequential-adaptive scheme, namely:

$$\left. \begin{aligned} \inf_{\rho, \{V_i\}, M_{\tilde{\theta}}} \widehat{\mathcal{C}} &= \inf_{\rho, \{V_i\}, M_{\tilde{\theta}}} \sup_{g \in G} \int d\tilde{\theta} \text{Tr}(\rho_g^N M_{\tilde{g}}) \mathcal{C}(g, \tilde{g}) \\ \inf_{\rho, \{V_i\}, M_{\tilde{\theta}}} \overline{\mathcal{C}} &= \inf_{\rho, \{V_i\}, M_{\tilde{\theta}}} \int d\tilde{\theta} \text{Tr}(\rho_g^N M_{\tilde{g}}) \mathcal{C}(g, \tilde{g}) \end{aligned} \right\} = \inf_{\rho \in \text{Lin}(\mathcal{H}_S^{\otimes N} \otimes \mathcal{H}_A), M_e} \int d\tilde{g} \text{Tr}(\rho(U_{\tilde{g}}^{\dagger \otimes N} \otimes \mathbb{1}) M_e(U_{\tilde{g}}^{\otimes N} \otimes \mathbb{1})) \mathcal{C}(e, \tilde{g}), \quad (5.44)$$

where in LHS $\rho_g^N = V_N \circ (U_g \otimes \mathbb{1}) \circ \dots V_1 \circ (U_g \otimes \mathbb{1}) \circ \rho$.

Proof. Consider an arbitrary quantum channel:

$$\mathcal{E} : \text{Lin}(\mathcal{H}_{\text{in}}) \rightarrow \text{Lin}(\mathcal{H}_{\text{out}}), \quad (5.45)$$

where in principle $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_{\text{out}}$ may correspond to the same physical system, yet in this formalism, they will be treated as separated Hilbert spaces. From Choi-Jemiołkowski isomorphism, the channel is uniquely defined by its Choi matrix:

$$E = \text{Choi}(\mathcal{E}) := \mathcal{E} \otimes \mathbb{1}(|\Omega\rangle\langle\Omega|) \quad (5.46)$$

where $|\Omega\rangle = \sum |i\rangle|i\rangle \in \mathcal{H}_{\text{in}} \otimes \mathcal{H}_{\text{in}}$. Indeed, by direct calculation, one may see:

$$\text{Tr}_{\text{in}}(E(\mathbb{1} \otimes \rho^T)) = \text{Tr}_{\text{in}}\left(\sum_{ij} \mathcal{E}(|i\rangle\langle j|) \otimes |i\rangle\langle j| \rho^T\right) = \sum_{ij} \mathcal{E}(|i\rangle\langle j|) \rho_{ji}^T = \mathcal{E}(\rho). \quad (5.47)$$

Moreover, to introduce consistent notation, also density matrix ρ itself as well as each POVM element M_i will be seen as quantum channels, respectively:

$$\begin{aligned} \rho : \mathbb{C} &\rightarrow \mathcal{H}_{\text{in}} : 1 \mapsto \rho, \\ M_i : \mathcal{H}_{\text{out}} &\rightarrow \mathbb{C} : \rho \mapsto \text{Tr}(M_i \rho). \end{aligned} \quad (5.48)$$

Especially, Choi matrix of ρ is simply ρ , while Choi matrix of M_k is its transposition, as

$$M_k \otimes \mathbb{1}(|\Omega\rangle\langle\Omega|) = \sum_{ij} \text{Tr}(M_k |i\rangle\langle j|) |i\rangle\langle j| = M^T. \quad (5.49)$$

By $V^{|\Omega\rangle}$ I will denote the Choi matrix of the unitary channel V , i.e. $V^{|\Omega\rangle} = (V \otimes \mathbb{1})|\Omega\rangle\langle\Omega|(V^\dagger \otimes \mathbb{1})$.

Next, let us derive the formula for the Choi matrix of the composition of the channels. Consider

now two channels, for which the output of the first one is the input of the second:

$$\begin{aligned}\mathcal{E}_1 &: \text{Lin}(\mathcal{H}_a) \rightarrow \text{Lin}(\mathcal{H}_b), \\ \mathcal{E}_2 &: \text{Lin}(\mathcal{H}_b) \rightarrow \text{Lin}(\mathcal{H}_c).\end{aligned}\tag{5.50}$$

By direct calculation, we derive the formula for the Choi matrix of their composition:

$$\begin{aligned}\mathcal{E}_2(\mathcal{E}_1(\rho)) &= \text{Tr}_b(E_2(\mathbb{1}_c \otimes [\text{Tr}_a(E_1(\mathbb{1}_b \otimes \rho^T))]^T)) = \\ &= \text{Tr}_a\left(\text{Tr}_b\left((E_2 \otimes \mathbb{1}_a) [(\mathbb{1}_c \otimes E_1)(\mathbb{1}_c \otimes \mathbb{1}_b \otimes \rho^T)]^T\right)\right) = \\ &= \text{Tr}_a\left(\underbrace{\text{Tr}_b\left((E_2 \otimes \mathbb{1}_a)(\mathbb{1}_c \otimes E_1^{T_b})\right)}_{E_{2,1}}(\mathbb{1}_c \otimes \rho^T)\right)\end{aligned}\tag{5.51}$$

where in the last step, we used the fact that trace over subspace Tr_a is invariant for partial transposition on this subspace, so $\text{Tr}_a([E_1(\mathbb{1}_b \otimes \rho^T)]^T) = \text{Tr}_a([E_1(\mathbb{1}_b \otimes \rho^T)]^T)^{T_a} = \text{Tr}_a([E_1(\mathbb{1}_a \otimes \rho^T)]^{T_b}) = \text{Tr}_a([E_1^{T_b}(\mathbb{1}_a \otimes \rho^T)])$.

We can go further: consider the channels with an arbitrary number of input and output Hilbert spaces, with the assumption that some of the output of $\mathcal{E}_1$ is the input of $\mathcal{E}_2$, for example:

$$\mathcal{E}_1 : \text{Lin}(\mathcal{H}_a \otimes \mathcal{H}_b) \rightarrow \text{Lin}(\mathcal{H}_c \otimes \mathcal{H}_d \otimes \mathcal{H}_f)\tag{5.52}$$

$$\mathcal{E}_2 : \text{Lin}(\mathcal{H}_f \otimes \mathcal{H}_g) \rightarrow \text{Lin}(\mathcal{H}_h)\tag{5.53}$$

Then Choi matrix of their composition $(\mathbb{1}_{c,d} \otimes \mathcal{E}_2) \circ (\mathcal{E}_1 \otimes \mathbb{1}_g)$ is given as:

$$E_{2,1} = \text{Tr}_f((\mathbb{1}_{c,d} \otimes E_2)(E_1 \otimes \mathbb{1}_g)^{T_f}).\tag{5.54}$$

which we call link product $E_2 * E_1$. Note that, as partial trace $\text{Tr}_f(\cdot)$ is invariant for partial transposition $\cdot^{T_f}$, the link product is commutative $E_2 * E_1 = E_1 * E_2$.

Note that if two channels do not have any common input-output, the Choi matrix of their composition is simply a tensor product of their Choi matrices $E_1 * E_2 = E_1 \otimes E_2$.

We are now ready to use this formalism to describe the adaptive scheme 2.3(c). Here $\mathcal{H}_{\text{in}} \simeq \mathcal{H}_{\text{out}} \simeq \mathcal{H}_S \otimes \mathcal{H}_A$. Moreover, we incorporate the N^{th} unitary gate V_N into the measurement to get a more

compact notation. The probability of output x is then given as:

$$\begin{aligned}
p(x|\boldsymbol{\theta}) &= \\
&= \text{Tr}(M_x((\mathcal{E} \otimes \mathbb{1}) \circ V_{N-1} \circ (\mathcal{E} \otimes \mathbb{1}) \circ \dots \circ V_{N-2} \dots (\mathcal{E} \otimes \mathbb{1}) \circ \rho)) \\
&= M_x^T * (E * |\Omega\rangle\langle\Omega|) * V_{N-1}^{|\Omega\rangle} * \dots * V_1^{|\Omega\rangle} * (E * |\Omega\rangle\langle\Omega|) * \rho \\
&\stackrel{(1)}{=} (E * |\Omega\rangle\langle\Omega| \dots * E * |\Omega\rangle\langle\Omega|) * (M_x^T * V_{N-1}^{|\Omega\rangle} * \dots * V_1^{|\Omega\rangle} * \rho) \\
&\stackrel{(2)}{=} (E \otimes |\Omega\rangle\langle\Omega|)^{\otimes N} * (M_x^T \otimes V_{N-1}^{|\Omega\rangle} \otimes \dots \otimes V_1^{|\Omega\rangle} \otimes \rho) \\
&\stackrel{(3)}{=} \text{Tr}((E \otimes |\Omega\rangle\langle\Omega|)^{\otimes N} (M_x^T \otimes V_{N-1}^{|\Omega\rangle} \otimes \dots \otimes V_1^{|\Omega\rangle} \otimes \rho)^T) \\
&= \text{Tr}((E \otimes |\Omega\rangle\langle\Omega|)^{\otimes N} (M_x \otimes V_{N-1}^{|\Omega\rangle T} \otimes \dots \otimes V_1^{|\Omega\rangle T} \otimes \rho^T)),
\end{aligned} \tag{5.55}$$

where in (1), I used the commutative of link product $*$, in (2), the fact that I link only channels with no common input-output space, while in (3) applied the composition formula.

Finally, we may perform the partial trace of all N ancillae to get a generalized Born rule for the probability of getting measurement result x :

$$p(x|\boldsymbol{\theta}) = \text{Tr}(DP_x), \quad \text{with} \quad D = E^{\otimes N}, \quad P_x = \text{Tr}_{\mathcal{H}_A^{\otimes N}} \left((\mathbb{1} \otimes |\Omega\rangle\langle\Omega|)^{\otimes N} (M_x \otimes \dots \otimes V_i^{|\Omega\rangle T} \dots \otimes \rho^T) \right), \tag{5.56}$$

where P_x is called a tester, including all information about the measurement strategy.

From definition of POVM, the set of testers $\{P_x\}$ satisfies $\sum_x P_x = \mathbb{1} \otimes \Xi^{(N)}$, while $\Xi^{(N)}$ is called normalization operator. These allow us for define $\tilde{P}_x, \tilde{D}$:

$$\begin{aligned}
P_x &= \left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right) \tilde{P}_x \left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right), \\
\tilde{D} &= \left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right) D \left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right),
\end{aligned} \tag{5.57}$$

such that $p(x) = \text{Tr}(\tilde{P}_x \tilde{D})$, where $\{\tilde{P}_x\}$ form POVM¹ and $\tilde{D}$ is the state (by construction is semi-positive and one-trace) on (in general abstract) Hilbert space $(\mathcal{H}_{\text{out}} \otimes \mathcal{H}_{\text{in}})^{\otimes N}$ (where $\dim(\mathcal{H}_{\text{in}}) = \dim(\mathcal{H}_{\text{out}}) = \dim(\mathcal{H}_S)$).

Now consider the case when the channel is given by unitary representation of the group $\mathcal{E}_g(\rho) = U_g \rho U_g^\dagger$, so $E_g = (U_g \otimes \mathbb{1}) |\Omega\rangle\langle\Omega| (U_g^\dagger \otimes \mathbb{1})$. In the same way, as in [Sec. 4.2](#), we justify that the optimal tester is a covariant one. Namely, for any tester $\{P'_g\}$ we may construct the corresponding

¹Note, that we cannot formally write $\tilde{P}_x = (\mathbb{1} \otimes \sqrt{\Xi^{(N)}})^{-1} P_x (\mathbb{1} \otimes \sqrt{\Xi^{(N)}})^{-1}$, as in general $\mathbb{1} \otimes \sqrt{\Xi^{(N)}}$ may has some zero eigenvalues. Still, as each $P_x \geq 0$, any zero-eigenvector of $\mathbb{1} \otimes \sqrt{\Xi^{(N)}}$ is simultaneously zero-eigenvector of all $P_x \geq 0$, so proper $\tilde{P}_x$ always exists. In the case where $\mathbb{1} \otimes \sqrt{\Xi^{(N)}}$ has some zero eigenvalues, formally, $\{\tilde{P}_x\}$ does not form a POVM on $(\mathcal{H}_{\text{out}} \otimes \mathcal{H}_{\text{in}})^{\otimes N}$, but it may be easily extended to it, adding as the last element the identity on $\ker(\mathbb{1} \otimes \sqrt{\Xi^{(N)}})$. That will not affect any further reasoning.

covariant one by proper averaging:

$$P_h = \mathcal{U}_h \left(\int_G dg \mathcal{U}_g^\dagger P'_g \mathcal{U}_g \right) \mathcal{U}_h^\dagger, \quad \text{with} \quad \mathcal{U}_h = (U_h \otimes \mathbb{1})^{\otimes N}, \quad (5.58)$$

where dg is the normalized Haar measure.

$$P_h = (U_h \otimes \mathbb{1})^{\otimes N} P_e (U_h^\dagger \otimes \mathbb{1})^{\otimes N}. \quad (5.59)$$

Next, we show that it may be realized in a parallel scheme with ancilla. Note that:

$$[\mathbb{1} \otimes \Xi^{(N)}, (U_h \otimes \mathbb{1})^{\otimes N}] = \left[\int_G h(\mu) d\mu P_h, (U_h \otimes \mathbb{1})^{\otimes N} \right] = 0 \quad (5.60)$$

Therefore the state $\tilde{D}$ may be written in the form:

$$\begin{aligned} \tilde{D} = & \left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right) (U_h \otimes \mathbb{1})^{\otimes N} |\Omega\rangle\langle\Omega| (U_h^\dagger \otimes \mathbb{1})^{\otimes N} \left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right) = \\ & (U_h \otimes \mathbb{1})^{\otimes N} \left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right) |\Omega\rangle\langle\Omega| \left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right) (U_h^\dagger \otimes \mathbb{1})^{\otimes N}, \end{aligned} \quad (5.61)$$

so the optimal cost may be obtained by using the input state $\left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right) |\Omega\rangle\langle\Omega| \left(\mathbb{1} \otimes \sqrt{\Xi^{(N)}} \right) \in (\mathcal{H}_S \otimes \mathcal{H}_A)^{\otimes N}$ (with $\dim(\mathcal{H}_S) = \dim(\mathcal{H}_A)$) in the parallel strategy [Fig. 2.3\(b\)](#), what was to be proven. $\square$

The natural question is if the above reasoning could be generalized for the noisy case. We can introduce the class of specific noisy channels for which the proof is still valid. Let us modify the unitary channel by adding classical fluctuation of the parameter itself, namely $\mathcal{E}_g(\rho) = \int dh p(h) U_{gh} \rho U_{gh}^\dagger$ with arbitrary probability distribution $p(h)$. That covers, for example, the common problem of the phase-shift estimation with the occurrence of dephasing noise. Note that adding such noise does not affect the crucial step of the proof [Eq. \(5.60\)](#), so indeed the statement is still valid. However, this is not a case for arbitrary noise, so the proof cannot be trivially extended for the general case.

CHAPTER 6

HEISENBERG LIMIT IN MULTIPARAMETER METROLOGY

6.1. Introduction

In this chapter, I will discuss the problem of the Heisenberg limit in multiparameter metrology, presenting the results from [Górecki and Demkowicz-Dobrzański [2022b,a]]. I will focus mainly on two aspects.

The first one is the potential advantage of estimating p parameters jointly, when compared to estimating them separately, spending in total the same amount of resources (i.e., usage of quantum gates). The problem has been extensively discussed in the literature (mainly within QFI formalism). However, there were some doubts about the interpretation of the results¹.

The central ambiguity concerns the issue of resource allocation between parameters when they are measured separately. It has been assumed in many works that, having n gates available in a single implementation, they should be immediately divided between p parameters [Humphreys et al. [2013a], Yuan [2016], Baumgratz and Datta [2016], Ying-Wang et al. [2018], Goldberg et al. [2020]]. In the presence of Heisenberg scaling, it would lead to:

$$\sum_{i=1}^p \Delta^2 \tilde{\theta}_i \propto p \times \frac{1}{k} \times \frac{1}{(n/p)^2} = \frac{p^3}{kn^2}. \quad (6.1)$$

Based on these assumptions, a significant advantage of the strategy, where the parameter are measured jointly, was postulated (typically of order $\sim p$).

However, as QCR bound is guaranteed to be saturable only in the limit $k \rightarrow \infty$, so to use this formalism, we need to assume large k . Therefore it is much more efficient to divide the number of repetition k between parameters, keeping n gates in every single trial [Yousefjani et al. [2017], Chen and Yuan [2019], Ho et al. [2020]]:

$$\sum_{i=1}^p \Delta^2 \tilde{\theta}_i \propto p \times \frac{1}{k/p} \times \frac{1}{n} = \frac{p^2}{kn^2}, \quad (6.2)$$

which leads to p times smaller final cost – only due to more reasonable resources division, without the necessity of measuring parameters jointly. That makes questionable previously postulated advantages of a joint strategy.

The highly nontrivial question is how the situation looks when only the total amount of resources

¹An alternative approach to quantifying the incompatibility between parameters without discussion the problem of dividing resources, has been proposed in [Ragy et al. [2016], Albarelli and Demkowicz-Dobrzanski [2021]]. However, it does not answer the problems I refer to, so it will not be discussed broader in this thesis.

$N = kn$ is restricted. Then, due to quadratic scaling with n and linear with k , the optimal choice would be $n = N, k = 1$ – but this cannot be fully analyzed within QFI formalism. In this chapter, I will discuss it in the minimax approach.

The second aspect I want to analyze is the advantage of a sequential adaptive scheme over a parallel one – again, I will compare the results obtained in both paradigms.

6.2. Asymptotic equivalence of the global and local minimax cost

I will compare the minimal cost obtainable within two paradigms. In the first one, all N resources (quantum gates) are used in a fully optimal way (which I will analyze within minimax formalism). In the second, there is a restriction that only n gates may be used in a single trial, while the experiment is repeated a large number of times k (which may be effectively described within QFI formalism).

While the first one demands defining a finite-sized set, in which the procedure is designed to work well Θ , the second is strictly focused on a single point θ_0 . To make a reasonable comparison, in a single parameter case in [Ch. 4](#), I have chosen $\Theta = [\theta_0 - \delta/2, \theta_0 + \delta/2]$ and I have written the exact bounds for finite region size δ , showing, that its impact decreases with increasing N .

Here I use an analogous idea, but I will not investigate the rate of convergence. Instead, I focus only on asymptotic results. A simple approach, which captures the essential problem, has been proposed in [\[Hayashi \[2011\]\]](#) (originally for single parameter case), so-called local asymptotic minimax cost. Let us define δ -neighborhood of θ_0 as $\Theta(\theta_0, \delta) = \{\theta : \forall_i \theta_i \in [\theta_{0i} - \delta/2, \theta_{0i} + \delta/2]\}$. Consider the sequence of triples: initial states, N unitary controls and measurements, for each possible N , namely $(\rho^N, \{V_i^N\}, \{M_{\tilde{\theta}}^N\})$. Then for N -gates protocol the output state is given as $\rho_{\theta}^N = V_N^N \circ (\mathcal{E}_{\theta} \otimes \mathbb{1}) \circ \dots \circ V_1^N \circ (\mathcal{E}_{\theta} \otimes \mathbb{1}) \circ \rho^N$. By local asymptotic minimax cost around point θ_0 with cost matrix C_{θ_0} we define:

$$\inf_{\{(\rho^N, \{V_i^N\}, \{M_{\tilde{\theta}}^N\})\}} \lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} N^2 \sup_{\theta \in \Theta(\theta_0, \delta)} \int d\tilde{\theta} \text{Tr}(\rho_{\theta}^N M_{\tilde{\theta}}^N) \text{tr}(C_{\theta_0}(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T), \quad (6.3)$$

Note that the order of taking limits $\lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty}$ is crucial here – for the opposite order, the trivial constant estimator $\tilde{\theta} = \theta_0$ would lead to zero cost, while the current form corresponds to the analysis from [Ch. 4](#), but without the necessity of performing calculations for finite δ/N . Now, I will generalize the reasoning from [\[Hayashi \[2011\]\]](#) (originally performed for $U(1)$ estimation) for our case.

Theorem. Given a channel $\mathcal{E}_g(\rho) = U_g \rho U_g^\dagger$, being a unitary representation of the group $g \in G$. Let $\theta = [\theta_1, \dots, \theta_p]$ be the local parametrization of the group element around some point g_{θ_0} . Let $\mathcal{C}(g_{\theta}, g_{\tilde{\theta}})$ be the cost invariant for acting the group, and C_{θ} is its Hessian with respect to variable $\tilde{\theta}$, at point θ (we also assume some basic regularity properties, which will be given later). Then,

for the most general adaptive scheme, the local asymptotic minimax cost is the same as the global asymptotic minimax cost (while the second one is a covariant problem):

$$\begin{aligned} \forall \theta_0 \quad & \inf_{\{(\rho^N, \{V_i^N\}, \{M_\theta^N\})\}} \lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} N^2 \sup_{\theta \in \Theta(\theta_0, \delta)} \int d\tilde{\theta} \text{Tr}(M_\theta^N \rho_\theta^N) \text{tr}(C_{\theta_0}(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T) \\ & = \lim_{N \rightarrow \infty} N^2 \left(\inf_{(\rho^N, \{V_i^N\}, \{M_{\tilde{g}}^N\})} \sup_{g \in G} \int d\tilde{g} \text{Tr}(\rho_g^N M_{\tilde{g}}^N) \mathcal{C}(g, \tilde{g}) \right) \end{aligned} \quad (6.4)$$

Proof. We start with the observation that the asymptotic minimax cost for $g \in G_\delta \subset G$ is for sure smaller or equal to the asymptotic minimax cost for $g \in G$:

$$\begin{aligned} & \inf_{\{(\rho^N, \{V_i^N\}, \{M_{\tilde{g}}^N\})\}} \lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} N^2 \sup_{g \in G_\delta} \int d\tilde{g} \text{Tr}(M_{N, \tilde{g}} \rho_{N, g}) \mathcal{C}(g, \tilde{g}) \\ & \leq \lim_{N \rightarrow \infty} N^2 \left(\inf_{(\rho^N, \{V_i^N\}, \{M_{\tilde{g}}^N\})} \sup_{g \in G} \int d\tilde{g} \text{Tr}(\rho_g^N M_{\tilde{g}}^N) \mathcal{C}(g, \tilde{g}) \right), \end{aligned} \quad (6.5)$$

where $G_\delta \subset G$ is δ -neighborhood of the group neutral element e . I start by showing that they are in fact equal.

Similarly, as in the single parameter case, I introduce the notation for the minimax cost with a finite δ, N :

$$\text{minimax}(G_\delta, N) = \inf_{(\rho^N, \{V_i^N\}, \{M_{\tilde{g}}^N\})} \sup_{g \in G_\delta} \int d\tilde{g} \text{Tr}(\rho_g^N M_{\tilde{g}}^N) \mathcal{C}(g, \tilde{g}). \quad (6.6)$$

In this notation, RHS of Eq. (6.5) is simply $\lim_{N \rightarrow \infty} N^2 \text{minimax}(G, N)$, while LHS may be bounded from below by $\lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} N^2 \text{minimax}(G_\delta, N)$ (as taking inf outside of limit may only increase the value of an objective function). It is also clear that

$$\lim_{N \rightarrow \infty} N^2 \text{minimax}(G_\delta, N) \leq \lim_{N \rightarrow \infty} N^2 \text{minimax}(G, N). \quad (6.7)$$

What remains to be proven is the weak inequality in the opposite direction:

$$\lim_{N \rightarrow \infty} N^2 \text{minimax}(G_\delta, N) \stackrel{?}{\geq} \lim_{N \rightarrow \infty} N^2 \text{minimax}(G, N). \quad (6.8)$$

I do it by the following reasoning.

Having in total N gates, at first, I perform $\sqrt{N}$ independent single-gate experiments to find an approximated value g_{est} (for example, by using ML estimator). From the central limit theorem, the probability that the true value does not belong to δ -size neighborhood of the indication of estimator $g \notin g_{\text{est}} G_\delta$ (where $g_{\text{est}} G_\delta$ is the set G_δ shifted by the action of g_{est}), decreases exponentially with a

number of measurements $p_{\text{err}}(\sqrt{N}) \propto e^{-\sqrt{N}}$.

Then I spend remaining $N - \sqrt{N}$ gates to perform optimal estimation strategy for $g \in g_{\text{est}}G_\delta$. Since such a two-step strategy is, in general, only suboptimal, I have:

$$\text{minimax}(G, N) \leq p_{\text{err}}(\sqrt{N})c_{\text{max}} + (1 - p_{\text{err}}(\sqrt{N}))\text{minimax}(g_{\text{est}}G_\delta, N - \sqrt{N}), \quad (6.9)$$

where $c_{\text{max}} = \max_{g, \tilde{g}} \mathcal{C}(g, \tilde{g})$. Moreover, due to the symmetry of the whole problem, the RHS does not depend on g_{est} . After application of $\lim_{N \rightarrow \infty} N^2 \cdot$ to the both sites, and the use of $\lim_{N \rightarrow \infty} \frac{(N - \sqrt{N})^2}{N^2} = 1$, we obtain:

$$\lim_{N \rightarrow \infty} N^2 \text{minimax}(G, N) \leq \lim_{N \rightarrow \infty} N^2 (G_\delta, N - \sqrt{N}) = \lim_{N \rightarrow \infty} N^2 (G_\delta, N), \quad (6.10)$$

which, together with Eq. (6.7) gives:

$$\lim_{N \rightarrow \infty} N^2 \text{minimax}(G_\delta, N) = \lim_{N \rightarrow \infty} N^2 \text{minimax}(G, N). \quad (6.11)$$

Moreover, even if it was formulated for G_δ around neutral element e , due to covariant properties of the problem, it remains valid also for any $g_{\theta_0}G_\delta$.

Now I will argue that replacing $\mathcal{C}(g_\theta, g_{\tilde{\theta}})$ by its Hessian in θ_0 does not change the minimal obtainable value. To show that we need to justify that for the protocol minimizing LHS of Eq. (6.4), the probability of the gross error $\tilde{\theta} - \theta$ is negligible. I do it in the following way.

For any protocols sequence $\{(\rho^N, \{V_i^N\}, \{M_{\tilde{\theta}}^N\})\}$, for any finite δ_f we introduce following correction. If the result of measurement satisfies $\tilde{\theta} \in \Theta(\theta_0, \delta_f)$, we leave it unchanged. However, if any $\tilde{\theta}_i \notin [\theta_{0i} - \delta/2, \theta_{0i} + \delta/2]$, we change these result from $\tilde{\theta}_i$ to θ_{0i} . For any sequence, such a change may only decrease the value of LHS of Eq. (6.4), so for the optimal one, it does not change. Therefore, for the protocols locally optimal around θ_0 (in the sense of Eq. (6.3)), for any finite δ_f , the probability of getting result outside of $\Theta(\theta_0, \delta_f)$ is negligible.

On the other hand, for sufficiently regular cost function:

$$\forall \epsilon \exists \delta_f \forall_{\theta, \tilde{\theta} \in \Theta(\theta_0, \delta_f)} |\mathcal{C}(g_\theta, g_{\tilde{\theta}}) - (\theta - \tilde{\theta})^T C_{\theta_0} (\theta - \tilde{\theta})| \leq \epsilon \cdot (\theta - \tilde{\theta})^T C_{\theta_0} (\theta - \tilde{\theta}) T = \epsilon \cdot \text{tr}(C_{\theta_0} (\theta - \tilde{\theta})(\theta - \tilde{\theta})^T) \quad (6.12)$$

and from that:

$$\forall \epsilon \exists \delta_f \frac{1}{1 + \epsilon} \mathcal{C}(\theta, \tilde{\theta}) \leq \text{tr}(C_{\theta_0} (\theta - \tilde{\theta})(\theta - \tilde{\theta})^T) \leq \frac{1}{1 - \epsilon} \mathcal{C}(\theta, \tilde{\theta}). \quad (6.13)$$

The reasoning may be repeated for any ϵ , which proves the statement. $\square$

However, from Sec. 5.3, we know that for the covariant problem, the parallel covariant one is optimal

among all (adaptive or not) protocols. We may finally state:

$$\begin{aligned} \forall_{\theta_0} \inf_{\{(\rho^N, \{V_i^N\}, \{M_{\tilde{\theta}}^N)\}\}} \lim_{\delta \rightarrow 0} \lim_{N \rightarrow \infty} N^2 \sup_{\theta \in \Theta_\delta} \int d\tilde{\theta} \text{Tr}(M_{\tilde{\theta}}^N \rho_{\theta}^N) \text{tr}(C(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T) \\ = \lim_{N \rightarrow \infty} N^2 \inf_{\rho^N, M_e^N} \int d\tilde{g} \text{Tr}((U_{\tilde{g}}^\dagger \otimes \mathbb{1})^{\otimes N} M_e(U_{\tilde{g}} \otimes \mathbb{1})^{\otimes N} \rho) \mathcal{C}(e, \tilde{g}), \end{aligned} \quad (6.14)$$

Further, I will be interested mainly in the examples where the Hessian is the identity matrix, so the figure of merit is the sum of MSEs $\text{tr}(C(\theta - \tilde{\theta})(\theta - \tilde{\theta})^T) = \sum_{i=1}^p \Delta^2 \tilde{\theta}_i$. I will analyze the optimal achievable precision for the cases where:

- all N resources are used optimally (analyzed within minimax formalism) or there is an additional constraint for the maximal amount of resources used in single trial n , which is repeated many times $k \rightarrow \infty$ (analyzed with the use of quantum Cramér-Rao bound).
- all parameters are measured jointly (JNT) in a fully optimal way, or from the very beginning, the resources are divided between parameters, which are measured separately (SEP).

These give four options, for which the sum of variances $\sum_{i=1}^p \Delta^2 \tilde{\theta}_i$ will be labeled as $\widehat{\Delta^2 \tilde{\theta}}_{\text{JNT}}$, $\Delta^2 \tilde{\theta}_{\text{JNT}}^{\text{CR}}$, $\widehat{\Delta^2 \tilde{\theta}}_{\text{SEP}}$, and $\Delta^2 \tilde{\theta}_{\text{SEP}}^{\text{CR}}$ respectively (see also Fig. 6.1 for a graphical explanation on the example). In a further discussion, I will be interested only in asymptotic values ($N \rightarrow \infty$ or $k \rightarrow \infty$), which will be denoted by $\simeq$.

6.3. Multiple-phase interferometer

The first example I would like to discuss is the multiarm interferometer, with p sensing arms (corresponding to single-photon eigenstates $|i\rangle$ with $i = 1, 2, \dots, p$) and single reference arm (corresponding to state $|0\rangle$), see Fig. 6.1. The single photon evolution is therefore given by:

$$U_{\theta} = e^{i \sum_{i=1}^p \theta_i |i\rangle \langle i|}. \quad (6.15)$$

Let me start with the situation where the number of atoms used in a single trial is n , while the whole procedure is repeated many times k . As all generators mutually commute and therefore $[U_{\theta}, U_{\theta'}] = 0$, from discussion in Sec. 5.2 without loss of generality, we may restrict to parallel strategies, as adaptive sequential ones offer no advantage in this case. Moreover, we may restrict to fully symmetric space, where the number of photons in each arm fully characterizes the state. We have, therefore:

$$|\psi_{\theta}^n\rangle = \sum_{\mathbf{m}: \sum m_i \leq n} c_{\mathbf{m}} e^{i \mathbf{m} \cdot \theta} |m_1, \dots, m_p\rangle \quad (6.16)$$

with $\mathbf{m} = [m_1, \dots, m_p]$, where remaining $n - \sum_{i=1}^p m_i$ atoms are located in reference arm.

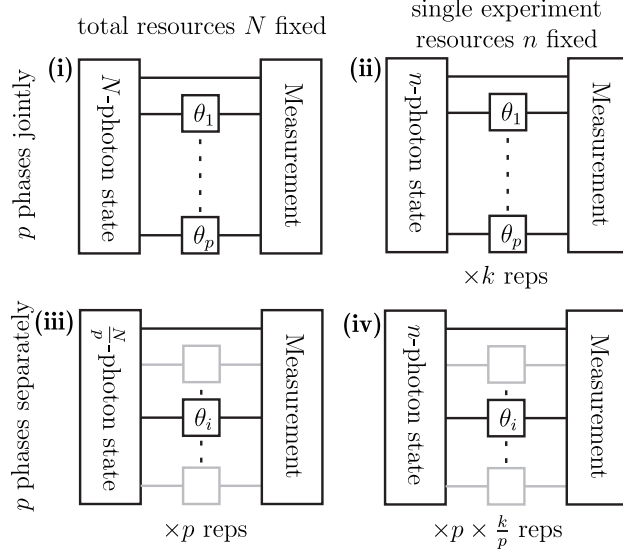


Figure 6.1: (Figure taken from [Górecki and Demkowicz-Dobrzański 2022a]) $p + 1$ -arm interferometer with single reference arm and p sensing arms is given. The phase shifts θ_i are estimated by putting as an input an arbitrary entangled state and performing the general measurement involving outputs of all arms. The problem is analyzed in four scenarios: the phases may be measured jointly or separately, where both cases are discussed with the constraint on all resources N or on resources used in a single trial n , where the trial is repeated k times. The sum of estimators MSE $\sum_{i=1}^p \Delta^2 \tilde{\theta}_i$ in each cases will be denoted as (i) $\widehat{\Delta^2 \tilde{\theta}}_{\text{JNT}}$, (ii) $\Delta^2 \tilde{\theta}_{\text{JNT}}^{\text{CR}}$, (iii) $\widehat{\Delta^2 \tilde{\theta}}_{\text{SEP}}$ and (iv) $\Delta^2 \tilde{\theta}_{\text{SEP}}^{\text{CR}}$.

In the case where one wants to measure all phases separately, the optimal strategy will use appropriate $n00n$ states (between reference 0^{th} and sensing i^{th} arms). In the limit of large k , resulting variance for each parameter would be $\frac{1}{k_i n^2}$, therefore, getting $k_i = k/p$ one end with:

$$\Delta^2 \tilde{\theta}_{\text{SEP}}^{\text{CR}} \simeq p \times \frac{1}{k/p} \times \frac{1}{n^2} = \frac{p^2}{kn^2}. \quad (6.17)$$

The optimal strategy of measuring all the phases jointly (within the paradigm of many repetitions) has been discussed in [Humphreys et al. 2013a]. The optimal state would be then:

$$|\psi_{\boldsymbol{\theta}}^n\rangle = \beta|0, \dots, 0\rangle + \alpha(e^{in\theta_1}|n, \dots, 0\rangle + \dots + e^{in\theta_p}|0, 0, \dots, n\rangle), \quad (6.18)$$

where coefficient minimizing the trace of the inverse of Fisher information are $\alpha = 1/\sqrt{p + \sqrt{p}}$ and $\beta = 1/\sqrt{1 + \sqrt{p}}$, which leads to the cost:

$$\Delta^2 \tilde{\theta}_{\text{JNT}}^{\text{CR}} \simeq \frac{1}{k} \text{Tr}(\mathbf{F}_Q^{-1}) = \frac{(1 + \sqrt{p})^2 p}{4kn^2} \stackrel{p \gg 1}{\approx} \frac{p^2}{4kn^2} \quad (6.19)$$

(CR bound is asymptotically saturable, as all SLDs mutually commute). With an increasing number of phases, the advantage of measuring them jointly converges to a constant factor $1/4$.

However, an interesting conjecture has been made by the Authors of [Humphreys et al. [2013a]] about the consequences of the above results for the situation when only the total amount of photons N is restricted, and one has full freedom in dividing them between n and k . They put forward the hypothesis that, as in the first approach, all p phases are measured independently, it would require (in total) p times more trials to saturate CR bound. Then, assuming that there exists some fixed minimal $k_{\min}$ needed for saturating CR, it would lead to $\Delta^2 \tilde{\theta}_{\text{SEP}}^{\text{CR}} \stackrel{?}{\simeq} p \times \frac{1}{k_{\min}} \times \frac{1}{(N/pk_{\min})^2} = \frac{p^3 k_{\min}}{N^2}$ and $\Delta^2 \tilde{\theta}_{\text{JNT}}^{\text{CR}} \stackrel{?}{\simeq} \frac{p^2}{4k_{\min}(N/k_{\min})^2} = \frac{p^2 k_{\min}}{4N^2}$. That would indicate the advantage, which scales linearly with p . This assumption is, however, unjustified, as CR bound is known only in the limit $k \rightarrow \infty$, and the rate of this convergence may depend on a particular model. Therefore, one cannot simply put the same number $k_{\min}$ to both equations to judge the advantage or its lack in terms of all resources used. Such discussion cannot be performed within QFI formalism at all.

To answer the question, I use minimax formalism. I am interested in asymptotic $N \rightarrow \infty$, so from Eq. (6.14) I may restrict to parallel strategies with covariant measurements. If the phases are measured separately, we use the solution from single parameter case Eq. (4.20), where only reference 0^{th} and sensing i^{th} arms are involved. As for each parameter, we spent N/p photons, it leads to:

$$\widehat{\Delta^2 \tilde{\theta}_{\text{SEP}}} \simeq p \times \frac{\pi^2}{(N/p)^2} = \frac{p^3 \pi^2}{N^3}, \quad (6.20)$$

so, comparing to Eq. (6.17), we observe different scaling with the number of parameters, as expected.

For the problem of measuring all phases jointly, the exact solution of the problem is beyond our reach. So instead, I derive a simple lower bound for achievable cost and then propose the state for which the cost is relatively close for the large N .

Consider N -photon input state optimal for measuring all phases jointly. Without loss of generality, we may assume that it is symmetric for replacing the sensing arms, so for any sensing arm, the expected number of photons will be $\leq N/p$. Therefore, we can bound from below the variance of every single phase using the bound derived for estimation with mean energy Eq. (4.73), having finally:

$$\widehat{\Delta^2 \tilde{\theta}_{\text{JNT}}} \gtrsim p \times \frac{4|A_0|^3}{27} \frac{1}{(N/p)^2} = \frac{4|A_0|^3}{27} \frac{p^3}{N^2} \approx \frac{1.89p^3}{N^2}. \quad (6.21)$$

Compared to Eq. (6.20), we see that the maximum potential advantage is again of order ~ 4 .

That it is – the fact that for this model, one cannot achieve better scaling with p by measuring all parameters jointly comes directly from the nature of the Heisenberg limit (understood in terms

of the average energy). It indicates that if one would like to think in terms of $k_{\min}$, in the joint strategy Eq. (6.18), it needs to be at least p times bigger than in a separate one.

To check if the bound is close to being saturable, for large N we use the same approximation as in the single-phase case. We replace $m_i \in \{0, 1, \dots, N\}$ with continuous ones $\frac{m_i}{N} \rightarrow \mu_i \in [0, 1]$ and characterize the state by p -dimensional wave function $\hat{f}(\boldsymbol{\mu})$ ²:

$$|\psi_{\boldsymbol{\theta}}^N\rangle = \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\boldsymbol{\mu} \hat{f}(\boldsymbol{\mu}) e^{iN\boldsymbol{\mu}\boldsymbol{\theta}} |\mu_1, \mu_2, \dots, \mu_p\rangle. \quad (6.22)$$

Applying covariant measurements:

$$|\tilde{\boldsymbol{\theta}}\rangle = \frac{1}{\sqrt{(2\pi/N)^p}} \int d\boldsymbol{\mu} e^{iN\boldsymbol{\mu}\tilde{\boldsymbol{\theta}}} |\boldsymbol{\mu}\rangle \quad (6.23)$$

and optimization over the initial state leads to:

$$\widehat{\Delta^2 \tilde{\boldsymbol{\theta}}_{\text{JNT}}} \simeq \min_f \int_{\mathbb{R}^p} d\tilde{\boldsymbol{\theta}} |\langle \tilde{\boldsymbol{\theta}} | \psi_{\boldsymbol{\theta}}^N \rangle|^2 (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^2 = \frac{1}{N^2} \min_f \int_{\mathbb{R}^p} d\tilde{\boldsymbol{\theta}} \left| f(\tilde{\boldsymbol{\theta}}) \right|^2 \tilde{\boldsymbol{\theta}}^2 \quad (6.24)$$

where we dropped the irrelevant dependence on $\boldsymbol{\theta}$. Going back to the $\boldsymbol{\mu}$ -representation, the minimization problem takes the following form:

$$\begin{aligned} \widehat{\Delta^2 \tilde{\boldsymbol{\theta}}_{\text{JNT}}} &\simeq \frac{1}{N^2} \min_f \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\boldsymbol{\mu} \hat{f}^*(\boldsymbol{\mu}) \left(\sum_{k=1}^p -\partial_{\mu_k}^2 \right) \hat{f}(\boldsymbol{\mu}), \\ &\text{with} \quad \int_{\forall \mu_i \geq 0, \sum_i \mu_i \leq 1} d\boldsymbol{\mu} |\hat{f}(\boldsymbol{\mu})|^2 = 1, \\ &\hat{f}(\boldsymbol{\mu}) = 0 \quad \text{for } \boldsymbol{\mu} \text{ on the boundary } (\mu_i = 0 \vee \sum_i \mu_i = 1). \end{aligned} \quad (6.25)$$

This makes the problem equivalent to finding the ground state of a free particle in the p -dimensional simplex-shape infinite potential well. While for some simplex shapes, the problem may be solved exactly by mapping to a model of one-dimensional p free fermions [Krishnamurthy et al. [1982], Turner [1984]], this is not our case (except $p \leq 2$ case [Li [1984]]). The exact solution is beyond our abilities. Still, knowing that the function needs to satisfy boundary consideration and (to minimize kinetic energy) it should be concave inside of the simplex, I propose a simple ansatz, for which the

²Note the change of convention $f \rightarrow \hat{f}$, compared to original paper [Górecki and Demkowicz-Dobrzański [2022a]], which has been applied here to state in consistency with notation from Ch. 4

energy may be calculated analytically:

$$\hat{f}(\boldsymbol{\mu}) \propto \left(\prod_{i=1}^p \mu_i \right)^\alpha \left(1 - \sum_{i=1}^p \mu_i \right)^\beta. \quad (6.26)$$

Direct calculations [Górecki and Demkowicz-Dobrzański [2022b]] shows, that for $\alpha = 3/2, \beta = \sqrt{p}$, the cost is equal:

$$\widehat{\Delta^2 \tilde{\boldsymbol{\theta}}}_{\text{JNT}} \simeq \frac{p(1 + 2\sqrt{p})^2 \sqrt{p}(4p + 2\sqrt{p} - 1)}{(8\sqrt{p} - 4)N^2} \stackrel{p \gg 1}{\approx} \frac{2p^3}{N^2}, \quad (6.27)$$

which is indeed close to Eq. (6.21) and offer constant (asymptotically independent on p) advantage over Eq. (6.20).

We see that even if the scaling of the total cost with a number of parameters in separate strategies is different in both paradigms, the advantage of the joint strategy over the separate one is quantitatively the same in both approaches. The advantage in both scenarios comes from the same reason. In joint protocol, the same reference arm is shared by all phases, and therefore more photons remain for sensing arms. As a result, in the limit of large p , the amount of atoms in the reference arm (for the optimal state) becomes negligible. Therefore this advantage needs to converge to constant and cannot improve the scaling with p .

6.4. Magnetic field with unknown direction – $SU(2)$ estimation

As the next example, I would like to discuss estimating three components of the magnetic field sensed by spin-1/2 particles. Similarly, as in the introductory example, I assume the elementary gate is obtained by the field acting sufficiently short in time such that the rotation angle $|\boldsymbol{\theta}| \leq \pi$, so the problem boils down to the estimation of the rotation. The single atom evolution is therefore given as:

$$U_{\boldsymbol{\theta}} = e^{i(\theta_1 \sigma_x + \theta_2 \sigma_y + \theta_3 \sigma_z)/2}. \quad (6.28)$$

Let me start again with the many repetitions scenario, with n gates in each of k trials.

Note that here, unlike in multiple-phase estimation problem, where each parameter could be measured regardless of the existence of the others, it is no longer the case. There is no subspace for which evolution would depend only on a single parameter. Therefore the values of the remaining two may affect estimation accuracy and should be treated as nuisance parameters, as discussed at the end of Sec. 5.2. We focus on estimating around the point $\boldsymbol{\theta}_0$. For simplicity of further formulas, let us choose the coordinates so that $\boldsymbol{\theta}_0 = [0, 0, \theta_0]$.

Instead of deriving the optimal cost by direct optimization, we bound it from below and then show the example saturating these bounds. As mentioned before, the variance of every single parameter

may be bounded by the inverse of the diagonal element of the QFI matrix:

$$k\Delta^2\tilde{\theta}_i \gtrsim [\mathbf{F}_Q^{-1}]_{ii} \geq ([\mathbf{F}_Q]_{ii})^{-1} \quad (6.29)$$

while $[\mathbf{F}_Q]_{ii}$ is simply single parameter θ_i QFI, so from Eq. (3.41) is equal:

$$[\mathbf{F}_Q]_{ii} = 4\Delta^2[-iU_{\boldsymbol{\theta}}^\dagger\partial_i U_{\boldsymbol{\theta}}] \quad (6.30)$$

which, after direct maximization of RHS over the state (independently for $i = 1, 2, 3$), gives

$$[\mathbf{F}_Q]_{11} \leq \text{sinc}^2(\theta_0), \quad [\mathbf{F}_Q]_{22} \leq \text{sinc}^2(\theta_0), \quad [\mathbf{F}_Q]_{33} \leq 1. \quad (6.31)$$

We see that the non-zero value of $|\boldsymbol{\theta}_0|$ decreases the value of QFI connected to the estimating components perpendicular to $\boldsymbol{\theta}$. It may be easily interpreted since as $\boldsymbol{\theta}$ is oriented in z direction, for all non-zero values of θ_0 the changes of z component are connected with changes in the length of the rotation vector, while changes of x, y - with changes of angle of rotation vector. Since $\theta_0 = 2\pi$ corresponds to full rotation, exactly at this point, the infinitesimal changes of the angle of the rotation vector do not change the channel at all. For direct calculations performed in spherical coordinates, see [Yuan [2016]].

For a single usage of the gate, the maximally entangled state $\frac{1}{\sqrt{2}}(|\uparrow\rangle_z|0\rangle + |\downarrow\rangle_z|1\rangle)$ saturates all three inequalities Eq. (6.31). At first sight, it may seem that this state distinguishes the z direction. However, all maximally entangled states of the qubit are mutually equivalent up to local transformations on the ancilla $\mathbb{1} \otimes V$ [Kraus [2010]] (which does not change the QFI of the output state, as commute with $U_{\boldsymbol{\theta}} \otimes \mathbb{1}$). Moreover, the resulting QFI matrix is diagonal, so $[\mathbf{F}_Q^{-1}]_{ii} = ([\mathbf{F}_Q]_{ii})^{-1}$. Together with the fact that SLDs commute on state, Eq. (6.29) is saturated, therefore for a single quantum gate (with k repetitions), we have:

$$\sum_{i=1}^3 \Delta^2\tilde{\theta}_i \simeq \frac{1}{k} \left(1 + \frac{2}{\text{sinc}^2(|\boldsymbol{\theta}|)} \right). \quad (6.32)$$

For n usage of gates in a single trial, the situation becomes a bit more complicated. The states optimal for measuring each of the parameters separately $\frac{1}{\sqrt{2}}(|\uparrow\rangle_x^{\otimes n}|0\rangle + |\downarrow\rangle_x^{\otimes n}|1\rangle)$, $\frac{1}{\sqrt{2}}(|\uparrow\rangle_y^{\otimes n}|0\rangle + |\downarrow\rangle_y^{\otimes n}|1\rangle)$, $\frac{1}{\sqrt{2}}(|\uparrow\rangle_z^{\otimes n}|0\rangle + |\downarrow\rangle_z^{\otimes n}|1\rangle)$ are no longer equivalent. For separate strategy, we therefore have:

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\text{SEP}}^{\text{CR}} \simeq \frac{1}{k/3} \frac{1}{n^2} + \frac{1}{k/3} \frac{1}{n^2 \text{sinc}^2(|\boldsymbol{\theta}|)} + \frac{1}{k/3} \frac{1}{n^2 \text{sinc}^2(|\boldsymbol{\theta}|)} = \frac{3}{kn^2} \left(1 + \frac{2}{\text{sinc}^2(|\boldsymbol{\theta}|)} \right). \quad (6.33)$$

To see if this result may be significantly improved by measuring all parameters jointly, note first that $\forall_{i \in \{x, y, z\}} [\mathbf{F}_Q]_{ii} \leq 4\Delta^2 J_i \leq 4\langle J_i^2 \rangle$, which further implies [Kolenderski and Demkowicz-Dobrzanski

[2008]]³:

$$\Delta^2 \tilde{\boldsymbol{\theta}}_{\text{JNT}}^{\text{CR}} \gtrsim \frac{1}{k} \text{Tr}(\mathbf{F}_Q^{-1}) \leq \frac{1}{k} \frac{9}{\text{Tr}(\mathbf{F}_Q)} \leq \frac{1}{k} \frac{9}{4(J_x^2 + J_y^2 + J_z^2)} \leq \frac{1}{k} \frac{9}{4 \cdot n/2(n/2 + 1)} = \frac{1}{k} \frac{9}{n(n+2)}. \quad (6.34)$$

For finite $|\boldsymbol{\theta}|$ the solution calculated directly gives [Baumgratz and Datta [2016]]⁴:

$$\Delta^2 \tilde{\boldsymbol{\theta}}_{\text{JNT}}^{\text{CR}} \simeq \frac{1}{k} \frac{3}{n(n+2)} \left(1 + \frac{2}{\text{sinc}^2(|\boldsymbol{\theta}|)} \right) \stackrel{|\boldsymbol{\theta}| \ll 1}{\approx} \frac{9}{kn(n+2)}. \quad (6.35)$$

Compared with Eq. (6.33), one can see that the advantage of the joint strategy over separate one $\frac{n+2}{n}$ disappears with an increasing number of gates used in a single trial.

However, looking at Eq. (6.32), one may easily notice that the parallel results may be significantly improved if sequential adaptive strategy with optimal unitary controls $\forall_i V_i = U_{\boldsymbol{\theta}}^\dagger$ will be applied. Then, for n usage of the gate, the QFI matrix is multiplied by n^2 when compared to single usage, as the output state is the same, while the derivative is multiplied by n ; that gives [Yuan [2016]]⁵

$$\sum_{i=1}^3 \Delta^2 \tilde{\theta}_i \simeq \frac{1}{k} \frac{1}{n^2} \left(1 + \frac{2}{\text{sinc}^2(|\boldsymbol{\theta}|)} \right) \stackrel{|\boldsymbol{\theta}| \ll 1}{\approx} \frac{3}{kn^2}. \quad (6.36)$$

We see that within the many repetitions scenario, the sequential adaptive scheme is needed to observe the asymptotic advantage of measuring all components jointly over measuring them separately.

Next, I would like to compare these results with the one obtained for the situation where all N gates are used optimally. To do that, I will use theorem Eq. (6.14) and the results obtained for the covariant problem. Note that for the covariant cost (invariant for the rotation), the exact form of the Hessian of this function, written in introduced coordinates x, y, z will depend on the values $\boldsymbol{\theta}$ (which compensates for the fact that also the optimal QFI matrix depends on $\boldsymbol{\theta}$). Therefore, to make the comparison reasonable and simple, I will choose the cost function for which $C = \mathbb{1}$ at $\boldsymbol{\theta} = 0$ and make the comparison around this point.

Note that in this case, the problem of measuring each of the three parameters separately is highly non-trivial, as, unlike in the multiple-phase example, there is no subspace of Hilbert space, for which the evolution would depend on the single parameter. As a result, even when wishing to measure one parameter, to construct a meaningful measurement and estimator, one needs to have some (at least approximate) knowledge of the values of the others.

It makes the problem of measuring the parameters separately a bit artificial, but in principle, we

³Note a missing 4 factor in Eq. (10) of [Kolenderski and Demkowicz-Dobrzanski [2008]].

⁴Note difference by factor 1/2 in the definition of the generators in Baumgratz and Datta [2016].

⁵Note difference by a factor 1/2 in the definition of the generators in [Yuan [2016]].

can formulate the problem where all N is divided between three independent experimenters (who cannot communicate) when each of them is focused on estimating a single component θ_i . Then, from the point of view of single parameter estimation, existing the others may only make the full problem harder (see next section for broader explanation), so the total cost for such strategy may be bounded from below analogously as in multiple-phase case:

$$\widehat{\Delta^2 \tilde{\theta}_{\text{SEP}}} \gtrsim 3 \times \frac{\pi^2}{(N/3)^2} = \frac{27\pi^2}{N^2} \quad (6.37)$$

(while here, there is no guarantee for saturating).

In the case where all parameters are measured jointly, from Eq. (6.14), we know that up to terms $o(1/N^2)$ the optimal strategy is the covariant one, realizable in the parallel scheme with ancilla, without the necessity of applying adaptivity. The problem has been solved in [Chiribella et al. [2004, 2005]], with a covariant cost:

$$e(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) = 6 - 2\text{Tr}(U_{\boldsymbol{\theta}}^{(1)} U_{\tilde{\boldsymbol{\theta}}}^{(1)\dagger}) \stackrel{|\boldsymbol{\theta}|, |\tilde{\boldsymbol{\theta}}| \ll 1}{\approx} 2 \sum_{i=1}^3 \Delta^2 \tilde{\theta}_i, \quad (6.38)$$

where $U_{\boldsymbol{\theta}}^{(1)}$ is a rotation matrix of a spin-1 particle. It was shown that the asymptotic minimal cost is $e(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \simeq 8\pi^2/N^2$, which is achievable using initial state:

$$|\psi_{in}\rangle = \sqrt{\frac{2}{N/2+1}} \sum_{j=0(\frac{1}{2})}^{N/2-1} \sin\left(\frac{(j+1)\pi}{J+1}\right) \left(\sum_{\alpha=1}^{2j+1} \frac{|j\alpha, m_j = \alpha\rangle}{\sqrt{2j+1}} \right), \quad (6.39)$$

where $|j\alpha, m_j = \alpha\rangle$ are states with a well defined total angular momentum j and its projection onto z direction, while α numerates different subspaces corresponding to equivalent irreducible representations of $SU(2)$. Note that here the degree of freedom connected with different irreducible representations plays the same roles as ancilla, so as a result, the optimal protocol does not need an additional one. After rescaling the cost function by factor $\frac{1}{2}$ we get

$$\widehat{\Delta^2 \tilde{\theta}_{\text{JNT}}} \simeq \frac{4\pi^2}{N^2}. \quad (6.40)$$

Note that these results in the asymptotic limit $N \rightarrow \infty$ may also be obtained using Fourier analysis [Hayashi, 2016, Section 12].

Compared to the previous multiple-phase example, we see significant qualitative differences. In the multiple-phase case, the relation between optimal separate and joint protocol within both paradigms was very similar; in all cases, adaptivity was not needed for optimal performance.

In contrast, for magnetic field estimation, there are significant differences between the results obtained in both paradigms. In the many repetitions scenario, the optimal parallel strategy offers asymptotically no advantage over separate protocol, while the adaptive one allows for a decrease in the cost by a factor 3. Contrarily, for the optimal usage of all resources N , already the parallel scheme offers a big advantage, which cannot be further improved by an adaptive one.

A similar analysis may be performed for the problem of estimating only two components of the magnetic field, where the third one is known to be zero, namely $U_{\boldsymbol{\theta}} = e^{i(\theta_1\sigma_x + \theta_2\sigma_y)}$, which in covariant formalism was analyzed in [Bagan et al. [2000, 2001]]. The results are qualitatively the same as in the example discussed above [Górecki and Demkowicz-Dobrzański [2022b]].

Looking at the two examples discussed above, one may ask if the discrepancy between the results obtainable in different paradigms is inextricably linked to the commutativity of the evolution generators.

From discussions performed in Ch. 5 we know that superiority of adaptive protocols over parallel ones is indeed strictly related to commutativity – if all generators mutually commute, the advantage cannot appear in any paradigm. Contrary, if generators do not commute, the superiority sequential-adaptive scheme may appear in the many repetitions scenario (while for optimal usage of all N resources, the entangled-parallel is guaranteed to be optimal).

What about the existing advantage of measuring parameters jointly vs. separately? In a moment, I will give an example to show that divergence on this matter between the two paradigms is **not** clearly related to the issue of commutation of operators. Before I do so, I will introduce some general theories.

6.5. General bound for precision in multiparameter unitary evolution

The problem of the multiparameter estimation with optimal usage of all N gates has known analytical solution for some specific group estimation problems: $SU(2)/U(1)$ [Bagan et al. [2000, 2001]], $SU(2)$ [Chiribella et al. [2004], Bagan et al. [2004], Chiribella et al. [2005], Hayashi [2006]], $SO(3)$ [Hayashi [2016]]; moreover, the existing of Heisenberg scaling with N has been proven for the general case of $SU(d)$ [Kahn [2007]]. However, there is a lack of the universal lower bound. Below I present the universal lower bound [Górecki and Demkowicz-Dobrzański [2022b]] (unfortunately not tight), being the generalization of Eq. (4.52).

Consider a unitary channel dependent on p parameters, which enter linearly via generators Λ_i , such that $U_{\boldsymbol{\theta}} = e^{i\sum_{i=1}^p \theta_i \Lambda_i}$. First, I want to justify that if one is interested in estimating only one of these parameters θ_i , the existence of another θ_j may only make the task more difficult and never easier. Namely, even in the presence of other generators, the bound Eq. (4.52) is still valid (but not necessarily saturable). For more compact notation, let me name by $\lambda[\Lambda]$ the difference of maximal

and minimal eigenvalues of operator Λ . Let me remind the bound [Eq. \(4.52\)](#) in this notation:

$$\widehat{\Delta^2 \tilde{\theta}} \geq \frac{\pi^2}{\lambda^2[\Lambda]N^2} \left(1 - \frac{8 \log(N\lambda[\Lambda]\delta)}{N\lambda[\Lambda]\delta} \right). \quad (6.41)$$

Next, note that a single gate may be arbitrarily well approximated as:

$$U_{\theta} = e^{i \sum_{i=1}^p \theta_i \Lambda_i} \stackrel{l \gg 1}{\approx} \left(e^{i \theta_i \frac{\Lambda_i}{l}} e^{i \sum_{j \neq i} \theta_j \frac{\Lambda_j}{l}} \right)^l \quad (6.42)$$

where from Trotter formula [[Cohen et al. \[1982\]](#)] the approximation is exact in the limit $l \rightarrow \infty$ (even if $[\Lambda_i, \Lambda_j] \neq 0$). Therefore N actions of the unitary U_{θ} may be seen as $l \cdot N$ action of $U_{\theta_i} = e^{i \theta_i \frac{\Lambda_i}{l}}$, with unitary controls $V = e^{i \sum_{j \neq i} \theta_j \frac{\Lambda_j}{l}}$ in between. As a result, the product $\lambda[\Lambda_i/l] \cdot (lN) = \lambda[\Lambda_i]N$ remains unchanged, so does the bound [Eq. \(6.41\)](#) (including the range of convergence). Therefore, for each parameter;

$$\widehat{\Delta^2 \tilde{\theta}_i} \gtrsim \frac{\pi^2}{\lambda^2[\Lambda_i]}. \quad (6.43)$$

To make use of the above bound in the multiparameter case, let me discuss the concept of re-parametrization.

For given unitary channel $U_{\theta} = e^{i \theta \cdot \Lambda}$ with cost matrix $\text{tr}(C(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T)$, for real invertible $p \times p$ matrix A , we may perform the transformation of parameters, generators and cost matrix:

$$\theta' = A^{-1} \theta, \quad \Lambda' = A^T \Lambda, \quad C' = A^T C A, \quad (6.44)$$

such that estimation problem remain unchanged:

$$\theta' \cdot \Lambda' = \theta \cdot \Lambda, \quad \text{tr}(C'(\tilde{\theta}' - \theta')(\tilde{\theta}' - \theta')^T) = \text{tr}(C(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T). \quad (6.45)$$

Therefore, even if the problem is initially formulated with θ' and arbitrary cost matrix C' , we may always reduce it to minimize the sum of variance by choosing $A = \sqrt{C}$, such that for θ the cost is simply $\sum_{i=1}^p \Delta^2 \tilde{\theta}_i$.

This allows us the bound from below the minimal achievable cost by assuming the most optimistic scenario, i.e., not only do the other parameters not interfere but there is a single initial state and a single optimum measurement for all parameters at the same time:

$$\widehat{\Delta^2 \tilde{\theta}_{\text{JNT}}} \gtrsim \sum_{i=1}^p \frac{\pi^2}{\lambda^2[\Lambda_i]}. \quad (6.46)$$

However, the parametrization for which the cost matrix is identity is not unique. Therefore, the

bound may be potentially improved by optimizing it over orthogonal transformation O :

$$\widehat{\Delta^2 \tilde{\boldsymbol{\theta}}_{\text{JNT}}} \geq \sum_{i=1}^p \max_O \frac{\pi^2}{\lambda^2[[O^T \boldsymbol{\Lambda}]_i]}. \quad (6.47)$$

Note that restricting to orthogonal transformations is crucial, as otherwise, we get cost matrix C with non-zero of diagonal elements, and then $\text{tr}(C(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})^T)$ cannot be bounded by the sum of Eq. (6.43).

For the problems discussed in Sec. 6.3, Sec. 6.4, the bound is rather far from being saturable. In multiple-phase case, by direct minimization over O , one can find that it leads to $\frac{5\pi^2}{2N^2}$ for $p = 2$ (which is slightly tighter than Eq. (6.21), but still two times smaller than an exact solution), while for $p \geq 3$ it is even weaker than Eq. (6.21). In the case of $SU(2)$, one can notice that application O does not improve the bound at all (as rotating Pauli matrices do not change their eigenvalues), which gives $\frac{3\pi^2}{N^2}$, which is a quarter less than the exact solution.

6.6. Magnetic field in two spatially separated points

At last, I would like to present a simple example for which the bound Eq. (6.47) will be tight, and also re-parametrization will play an important role. Consider a magnetic field oriented in z direction in two spatially separated points, sensed by spin-1/2 atoms. The Hilbert space of a single atom is spanned by vector $|i, \pm\rangle$, where coordinate $i = 1, 2$ is the spacial one. We have:

$$\theta_1 \Lambda_1 + \theta_2 \Lambda_2 = \theta_1 |1\rangle\langle 1| \otimes \sigma_z + \theta_2 |2\rangle\langle 2| \otimes \sigma_z, \quad (6.48)$$

so $\lambda_1 = \lambda_2 = 1$, what leads to the bound $\widehat{\Delta^2 \tilde{\boldsymbol{\theta}}_{\text{JNT}}} \gtrsim \frac{2\pi^2}{N^2}$.

To make the bound tighter, let us consider reparametrization: $\theta'_1 = (\theta_1 + \theta_2)/\sqrt{2}$, $\theta'_2 = (\theta_1 - \theta_2)/\sqrt{2}$, for which:

$$\begin{bmatrix} \theta_1 & 0 & 0 & 0 \\ 0 & -\theta_1 & 0 & 0 \\ 0 & 0 & \theta_2 & 0 \\ 0 & 0 & 0 & -\theta_2 \end{bmatrix} = \begin{bmatrix} (\theta'_1 + \theta'_2)/\sqrt{2} & 0 & 0 & 0 \\ 0 & -(\theta'_1 + \theta'_2)/\sqrt{2} & 0 & 0 \\ 0 & 0 & (\theta'_1 - \theta'_2)/\sqrt{2} & 0 \\ 0 & 0 & 0 & -(\theta'_1 - \theta'_2)/\sqrt{2} \end{bmatrix} \quad (6.49)$$

so $\lambda_1 = \lambda_2 = \frac{1}{\sqrt{2}}$ which gives $\widehat{\Delta^2 \tilde{\boldsymbol{\theta}}_{\text{JNT}}} \gtrsim \frac{4\pi^2}{N^2}$. Moreover, this bound turns out to be achievable.

To see that, let as reorder the eigenbasis to be $\{|1, +\rangle, |2, -\rangle, |2, +\rangle, |1, -\rangle\}$. Then the matrix Eq. (6.49) takes the specific form:

$$\theta'_1 \Lambda'_1 + \theta'_2 \Lambda'_2 = \tilde{\theta}_1 \frac{1}{\sqrt{2}} \mathbb{1} \otimes \sigma_z + \tilde{\theta}_2 \frac{1}{\sqrt{2}} \sigma_z \otimes \mathbb{1} \quad (6.50)$$

Then, by renaming $\{|1, +\rangle, |2, -\rangle, |2, +\rangle, |1, -\rangle\} \rightarrow \{|+, +\rangle, |+, -\rangle, |-, +\rangle, |-, -\rangle\}$, on the mathematical level we can identify the Hilbert space of single particle with a 2-dimensional spacial degree of freedom with Hilbert space of 2 independent spin-1/2 particles (with no spacial degree of freedom).

It makes it clear that, therefore, $\tilde{\theta}_1, \tilde{\theta}_2$ may be simultaneously measured optimally with the same input state without interfering with each other. More precisely, we can write the Hilbert space as $\mathcal{H} = \mathbb{C}^{\otimes 2} \otimes \mathbb{C}^{\otimes 2}$, so for N copies in the parallel strategy, we would have $\mathcal{H}^{\otimes N} = (\mathbb{C}^{\otimes 2} \otimes \mathbb{C}^{\otimes 2})^{\otimes N} = (\mathbb{C}^{\otimes 2})^{\otimes N} \otimes (\mathbb{C}^{\otimes 2})^{\otimes N}$. Therefore the state of the form $|\text{sin}\rangle \otimes |\text{sin}\rangle$ (where $|\text{sin}\rangle$ is the one of the form Eq. (4.20)) with covariant measurement will saturate the bound:

$$\widehat{\Delta^2 \tilde{\theta}}_{\text{JNT}} \simeq \frac{4\pi^2}{N^2}. \quad (6.51)$$

We see that such a strategy offers the advantage over the one where θ_1, θ_2 are measured separately, which leads to:

$$\widehat{\Delta^2 \tilde{\theta}}_{\text{SEP}} \simeq 2 \times \frac{\pi^2}{(N/2)^2} = \frac{8\pi^2}{N^2}. \quad (6.52)$$

An interesting observation may be concluded if one looks at the problem from a many-repetition scenario perspective. Then for joint strategy, by the same reasoning, one obtains

$$\Delta^2 \tilde{\theta}_{\text{JNT}}^{\text{CR}} \simeq \frac{4}{kn^2}. \quad (6.53)$$

However, for measuring parameters separately:

$$\Delta^2 \tilde{\theta}_{\text{SEP}}^{\text{CR}} \simeq 2 \times \frac{1}{k/2} \times \frac{1}{n^2} = \frac{4}{kn^2}, \quad (6.54)$$

so, in this case, there is no advantage to applying a more advanced joint strategy! Unlike in previously discussed examples, we observe qualitative differences about existing advantages depending on the paradigm.

For the number of spatial points p bigger than 2, the above construction cannot be performed, as, simply speaking, we lose the identity $2p = 2^p$. The broad discussion about the case $p > 2$ may be found in [Górecki and Demkowicz-Dobrzański [2022b]].

One more feature should be mentioned here. In all discussed examples in the separate strategy, we measured the initially introduced parameters. Especially in the last example, we were focused on measuring θ_1, θ_2 , while in principle, one could consider as a separate strategy measuring independently θ'_1, θ'_2 , or, in general, any linear combinations of them. From a practical point of view, typically, some parametrization is more natural, and in trying to measure their linear combination, we simply lose the simplicity, which is the crucial motivation for a separate strategy (here: restrict-

ing the measurements local in space). Still, it is a reasonable comment from a general point of view. However, it turns out that for all discussed examples, this "natural" parametrization was indeed the optimal one for measuring parameters separately, as shown in [App. C](#).

6.7. Summary

The analysis of the examples shows that, in contrast to the estimation of a single unitary parameter, in the multiparameter case, there is no clear and unambiguous correspondence between the results obtained in the many repetitions paradigm (analyzed within CR formalism) and those corresponding to the optimal use of all resources (analyzed using MM). I analyzed in detail two issues

- superiority of sequential-adaptive scheme over the engaged-parallel one.
- the advantage of measuring all parameters jointly, compared to the strategy, where they are measured separately.

In the first case, the existence of this superiority is strictly related to the commutativity of evolution generators, which is consistence with theorems from [Ch. 5](#). If all generators mutually commute, the optimal results may be obtained within a parallel scheme (in both paradigms). However, in non-commuting cases, in CR adaptive scheme may lead to significantly better results, while in MM a parallel scheme is asymptotically sufficient to obtain the optimal result (which I illustrated by the problem of estimation of the three components of the magnetic field).

On the other hand, the divergence of the conclusions obtained in the two paradigms regarding the gain from measuring the parameters jointly is not related to commutativity. I have shown an example of estimating one component of the magnetic field in two spatially separated points (so generators mutually commute), where CR postulates do advantage from measuring parameters jointly, while the MM approach allows for decreasing cost by a factor of 2.

In general, the conclusions about the usefulness of carrying out a joint measurement made solely with CR may be overly pessimistic. For optimal use of all resources (MM), the advantage may exist for models that do not exist in a many repetitions scenario (CR).

CHAPTER 7

CONCLUSION AND OUTLOOK

For single-parameter unitary noiseless estimation, I have shown that the fundamental bound on the accuracy of the estimation with a limited number of the resources N is π times larger than would result from naive use of the error propagation formula or exploiting the quantum Fisher information (without considering multiple repetitions). The result considers any evolution generator (including one with a continuous spectrum) and remains valid for arbitrary entangled input states and a more general sequential adaptive scheme. The bound is asymptotically saturable. The rate of convergence depends on the ratio between the generator's spectral width and the size of the region of the parameter's occurrence. The superiority of the adaptive sequential scheme over the parallel one asymptotically disappears with increasing N . An analogous analysis of converging to the fundamental bound was performed (only for the parallel strategy) when only the average number of resources is limited (in this case, we are not discussing an adaptive strategy). In both cases, in the limit of $N \rightarrow \infty$, up to the leading order, the problems can be equated with estimating a completely unknown phase shift in a two-arm interferometer.

The analysis of the analogous problem in the presence of noise remains an open question. On the one hand, the QFI formalism offers extensive quantum error-correction tools and protocols, offering hope for recovering Heisenberg scaling. Unfortunately, their successful implementation requires almost exact knowledge of the parameter value from the very beginning. An actual implementation requires sequential tweaking of the circuit settings, considering the knowledge gained so far. There has been no thorough analysis of whether this would allow the MSE to keep scaling as $1/N^2$.

In the multi-parameter case, I compared comparing optimal metrological protocols for cases when the experiment is repeated many times (analyzed within CR formalism) or when all available resources can be accumulated together (analyzed within MM formalism). In the case of the multi-arm interferometer, the results in both cases are similar. However, in the case of the estimation of the three components of the magnetic field, the results are qualitatively different – CR predicts an asymptotic lack of benefit from measuring parameters jointly in a parallel scheme (however, it postulates a substantial advantage of the sequential adaptive scheme). In contrast, in the MM analysis, a substantial gain is already apparent for the parallel scheme (which cannot be improved by adaptiveness). It is known that the divergence in the utility of adaptivity in the two paradigms occurs only when the evolution generators do not commute. However, further analysis shows that the divergence in the benefit of measuring the parameters jointly (versus measuring them separately) is unrelated to the commuting of the generators. I also derived a simple (not necessarily saturated) lower bound on the achievable MSE. The problem of deriving a general saturable bound for multi-parameter metrology (even without noise) remains open.

APPENDIX A

EXAMPLE: INTERFEROMETER WITH LOSSES

Consider two arms interferometer [Fig. 2.2](#) with symmetric photon losses, i.e. for each arm only $\eta < 1$ of the signal go further, while $1 - \eta$ part is loss. For such a model the bound [Eq. \(3.61\)](#) gives [\[Demkowicz-Dobrzański and Maccone, 2014, Supplemental Material, Erasure\]](#):

$$\min_{\{K_k\}, \beta=0} \|\alpha\| = \frac{\eta}{4(1-\eta)} \implies \Delta^2 \tilde{\theta} \geq \frac{1}{kF_Q} \geq \frac{1}{kn} \frac{1-\eta}{\eta}, \quad (\text{A.1})$$

where n is the number of photons used in a single iteration.

It turns out that this bound may be saturated by putting to the upper arm the coherent state with the average value of photons equal $\bar{n}$ and the vacuum states weakly squeezed by a factor e^{-2r} to the lower one [\[Caves \[1981\]\]](#), followed by simply estimating the value of phaseshift from the difference of photon counting in both counter¹.

The expectation value of this difference is then

$$\langle N_- \rangle = \frac{\eta}{2} \cos(\theta) (\bar{n} - \sinh^2 r), \quad (\text{A.2})$$

so, unlike in the case of $n00n$ states or similar ones, the range of effectiveness of such strategy is finite and does not shrink with increasing $\bar{n}$. From the error propagation formula we have:

$$\Delta^2 \tilde{\theta} = \frac{\langle \Delta^2 N_- \rangle}{\left| \frac{d\langle N_- \rangle}{d\theta} \right|^2}, \quad (\text{A.3})$$

with

$$\left| \frac{d\langle N_- \rangle}{d\theta} \right|^2 = \frac{\eta^2}{4} \sin^2(\theta) (\bar{n} - \sinh^2(r))^2 \quad (\text{A.4})$$

$$\begin{aligned} \langle \Delta^2 N_- \rangle = \frac{\eta^2}{4} & \left(\cos^2(\theta) (\bar{n} + \frac{1}{2} \sinh^2(2r)) + \right. \\ & \left. \sin^2(\theta) (\bar{n} (\cos^2 \phi e^{-2r} + \sin^2 \phi e^{+2r}) + \sinh^2 r) + \frac{1-\eta}{\eta} (\bar{n} + \sinh^2 r) \right) \end{aligned} \quad (\text{A.5})$$

where ϕ is the angle between the phase of coherent input and squeezing direction, which should be set $\phi = 0$ for optimal performance. In asymptotic limit, choosing r such as $\frac{\eta}{1-\eta} \ll \sinh^2 r \ll \bar{n}$ we

¹Strictly speaking, as this strategy involves an indefinite number of photons with fixed mean value $\bar{n}$, the bound [Eq. \(A.1\)](#) does not apply here directly (as it was derived with the assumption of a well-defined number of gates used). Still, as we consider the case when the variance of the number of photons is much smaller than the square of its mean number, we do not expect any significant effects (contrary to the ones from [Sec. 3.6](#)) coming from this fact

may approximate above by:

$$\Delta^2 \tilde{\theta} \approx \frac{\cos^2 \theta + \frac{1-\eta}{\eta}}{\sin^2 \theta \bar{n}}. \quad (\text{A.6})$$

Clearly, the optimal point is at $\theta = \pi/2$, when it saturates the bound [Eq. \(A.1\)](#). However, the MSE does not change significantly for $\cos^2(\theta) \ll \frac{1-\eta}{\eta}$, so indeed, the strategy works well in a finite-sized region.

Such a method has been used in the LIGO gravitational wave detector [[LIGO \[2011\]](#), [Aasi et al. \[2013\]](#)]. In practice the procedure implementing squeezed vacuum to the lower arm significantly increases the noise, so the condition $\frac{\eta}{1-\eta} \ll \sinh^2 r$ may be not satisfied. Moreover, to talk about quantum advantage, one would compare the resulting variance with the one which would be obtained for the coherent light without these additional losses or – equivalently – for coherent light with the same **output energy** (which is in practice true limitation, given by photons counters abilities). In a small neighborhood of $\theta = \pi/2$ the formula for MSE may be approximated as:

$$\Delta^2 \tilde{\theta} \approx \frac{\eta e^{-2r} + (1 - \eta)}{\eta \bar{n}}. \quad (\text{A.7})$$

Therefore, the measure of quantum advantage is the value of the nominator in the above equation. In [[LIGO \[2011\]](#)] for squeezing rate $e^{-2r} = 0.1$ (10dB) and detection efficiency $\eta = 0.62$ decreasing the MSE by a factor 0.47 (3.5dB) has been observed, in accordance with the above formula.

In [[LIGO \[2011\]](#)] for squeezing rate $e^{-2r} = 0.93$ (10.3dB) and detection efficiency $\eta = 0.44$ was obtained, which, according to the above formula, should lead to decreasing MSE by 0.6 (2.2dB). The observed effect was minimally weaker 2.15dB, which has been identified as the result of normal fluctuation of the phase ϕ of order 37 ± 6 mrad.

APPENDIX B

DERIVATION OF [Eq. \(4.50\)](#)

Consider probability distribution

$$p_{\alpha,L}(\theta) = \mathcal{N}_\alpha L \text{sinc}^4 \left(\pi \alpha \sqrt{(L\theta/4\alpha)^2 - 1} \right) = \mathcal{N}_\alpha L \frac{\sinh^4 \left(\pi \alpha \sqrt{1 - (L\theta/4\alpha)^2} \right)}{\left(\pi \alpha \sqrt{1 - (L\theta/4\alpha)^2} \right)^4}, \quad (\text{B.1})$$

where $L = 2N_0$ is the bandwidth, α determines the shape and $\mathcal{N}_\alpha$ may be bounded by:

$$\mathcal{N}_\alpha \lesssim 4\sqrt{2}\pi^4 \alpha^{7/2} e^{-4\pi\alpha}, \quad (\text{B.2})$$

where the bound is tight for big α [[Górecki et al. \[2020\]](#)]. As shown in [[Górecki et al. \[2020\]](#)] only exponentially small (with α) part lays outside of region $[-4\alpha/L, 4\alpha/L]$ and, therefore, for our purpose we choose $\delta/2 = 4\alpha/L$.

For such a function, I will show that for properly chosen L , inequality [Eq. \(4.50\)](#)

$$\frac{\pi^2}{(\lambda N + L/2)^2} - R_2 \stackrel{?}{\geq} \frac{\pi^2}{\lambda^2 N^2} \left(1 - \frac{8 \log(N\lambda\delta)}{N\lambda\delta} \right) \quad (\text{B.3})$$

holds. Using general identity $\frac{1}{(1+x)^2} = 1 - 2x + \frac{3x^2+x^3}{(1+x)^2}$ we may rewrite LHS of [Eq. \(B.3\)](#) as:

$$\frac{\pi^2}{(\lambda N + L/2)^2} - R_2 = \underbrace{\frac{\pi^2}{\lambda^2 N^2} \left(1 - \frac{L}{\lambda N} \right)}_{B_1} + \underbrace{\frac{\pi^2}{\lambda^2 N^2} \frac{3(L/2\lambda N)^2 + 2(L/2\lambda N)^3}{(1 + L/2\lambda N)^2}}_{B_2} - R_2. \quad (\text{B.4})$$

Next, I will show that for proper L , B_1 is equal to RHS of [Eq. \(B.3\)](#), while B_2 is strictly positive for big $N\lambda\delta$.

First I bound from above R_2 :

$$\begin{aligned} R_2 &= 2\mathcal{N}_\alpha \int_{4\alpha/L}^{+\infty} d\theta L \text{sinc}^4 \left(\pi \alpha \sqrt{(L\theta/4\alpha)^2 - 1} \right) (\theta + \delta/2)^2 \\ &= 2\mathcal{N}_\alpha \frac{4\alpha}{L} \int_1^{+\infty} dx L \text{sinc}^4 \left(\pi \alpha \sqrt{x^2 - 1} \right) (x+1)^2 \left(\frac{4\alpha}{L} \right)^2 \leq \\ &\leq 2\mathcal{N}_\alpha L \left(\frac{4\alpha}{L} \right)^3 \left(\int_1^2 dx (x+1)^2 + \frac{1}{\pi^4 \alpha^4} \int_2^{+\infty} dx \frac{(x+1)^2}{(x^2-1)^2} \right) \\ &= 2\mathcal{N}_\alpha L \left(\frac{4\alpha}{L} \right)^3 \left(\frac{19}{3} + \frac{1}{\pi^4 \alpha^4} \right)^{\alpha > 1/2} \leq 14\mathcal{N}_\alpha L \left(\frac{4\alpha}{L} \right)^3, \quad (\text{B.5}) \end{aligned}$$

where inequalities $\text{sinc}(x) \leq 1$ and $\text{sinc}(x) \leq 1/x$ were used.

Now, choosing $L = \frac{8}{\delta} \log(N\delta\lambda)$ (so $\alpha = \log(N\delta\lambda)$) we obtain:

$$B_1 = \frac{\pi^2}{\lambda^2 N^2} \left(1 - \frac{8 \log(N\lambda\delta)}{N\lambda\delta} \right) \quad (\text{B.6})$$

and, introducing $y = N\delta\lambda$ and $z(y) = 4 \log(y)/y$ for more compact notation,

$$L^2 \cdot B_2 = \pi^2 (2z)^2 \frac{3z^2 + 2z^3}{(1+z)^2} - 14 \mathcal{N}_{\log(y)} (4 \log(y))^3. \quad (\text{B.7})$$

By numerical calculation, it may be shown that the above is positive for $y \geq 2$, which was to be shown.

APPENDIX C

BROADER DISCUSSION – REPARAMETRIZATION

Here I recall the reasoning from [Górecki and Demkowicz-Dobrzański 2022b]. Let $\boldsymbol{\theta} = [\theta_1, \dots, \theta_p]^T$ be the parametrization for which the cost matrix is identity $C = \mathbb{1}$. Consider the extended procedure of measuring parameters separately with usage totally N gates, SEP+. Namely:

- first apply the linear transformation $\boldsymbol{\theta}' = A^{-1}\boldsymbol{\theta}$.
- then estimate each of θ'_i , using N_i gates, such that $\sum_{i=1}^p N_i = N$.

From resulting $\tilde{\boldsymbol{\theta}}'$ reconstruct $\tilde{\boldsymbol{\theta}} = A\tilde{\boldsymbol{\theta}}'$. We want to minimize $\sum_{i=1}^p \Delta^2 \tilde{\theta}_i$ over transformation A , resources' distribution $\sum_{i=1}^p N_i = N$ and single sequential-adaptive protocol for each $\tilde{\theta}'_i$. I will derive a simple lower bound for the final precision obtainable in such a procedure.

Note that if all parameters θ_i are measured separately with the usage of different probes, the resulting covariance matrix will be diagonal (as indications of different estimators come from uncorrelated measurements results):

$$\Sigma' = \text{diag}(\Delta^2 \tilde{\theta}'_1, \dots, \Delta^2 \tilde{\theta}'_p). \quad (\text{C.1})$$

Therefore, the final cost, which is to be minimized may be written as follows:

$$\sum_{i=1}^p \Delta^2 \tilde{\theta}_i = \text{Tr}(\Sigma) = \text{Tr}(AA^T \Sigma') = \sum_{i=1}^p [AA^T]_{ii} \Delta^2 \tilde{\theta}_i. \quad (\text{C.2})$$

Using Eq. (6.43) we may write:

$$\widehat{\Delta^2 \tilde{\boldsymbol{\theta}}}_{\text{SEP+}} \gtrsim \min_{A, N_j} \sum_{j=1}^p \frac{\pi^2}{N_j^2} \frac{[A^T A]_{jj}}{\lambda^2([A^T \boldsymbol{\Lambda}]_j)} \geq \min_{A, N_j} \sum_{j=1}^p \frac{\pi^2}{N_j^2} \left(\min_i \frac{[A^T A]_{ii}}{\lambda^2([A^T \boldsymbol{\Lambda}]_i)} \right) = \frac{p^3 \pi^2}{N^2} \min_{A, i} \frac{[A^T A]_{ii}}{\lambda^2([A^T \boldsymbol{\Lambda}]_i)}. \quad (\text{C.3})$$

For any fixed i , RHS depends on i^{th} column of A , the minimization is equivalent to minimization over single vector $\mathbf{a}$:

$$\min_{A, i} \frac{[A^T A]_{ii}}{\lambda^2([A^T \boldsymbol{\Lambda}]_i)} = \min_{\mathbf{a}} \frac{|\mathbf{a}|^2}{\lambda^2[\mathbf{a}\boldsymbol{\Lambda}]} = \min_{\mathbf{a}: |\mathbf{a}|^2=1} \frac{1}{\lambda^2[\mathbf{a}\boldsymbol{\Lambda}]} = \frac{1}{\max_{\mathbf{a}: |\mathbf{a}|^2=1} \lambda^2[\mathbf{a}\boldsymbol{\Lambda}]}, \quad (\text{C.4})$$

so we get:

$$\widehat{\Delta^2 \tilde{\boldsymbol{\theta}}}_{\text{SEP+}} \gtrsim \frac{p^3 \pi^2}{N^2} \frac{1}{\max_{\mathbf{a}: |\mathbf{a}|^2=1} \lambda^2[\mathbf{a}\boldsymbol{\Lambda}]}. \quad (\text{C.5})$$

Similarly, in the many repetitions scenario:

$$\Delta^2 \tilde{\boldsymbol{\theta}}_{\text{SEP}+}^{\text{CR}} \gtrsim \frac{p^2}{kn^2} \frac{1}{\max_{\boldsymbol{a}: |\boldsymbol{a}|^2=1} \lambda^2[\boldsymbol{a}\boldsymbol{\Lambda}]}.$$
(C.6)

After applying the examples discussed in [Ch. 6](#), it shows, that the parametrization, which was used, was indeed the one minimizing the cost obtainable in a separate strategy.

BIBLIOGRAPHY

- Junaid Aasi, Joan Abadie, BP Abbott, Richard Abbott, TD Abbott, MR Abernathy, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Enhanced sensitivity of the ligo gravitational wave detector by using squeezed states of light. *Nature Photonics*, 7(8):613–619, 2013. doi: <https://doi.org/10.1038/nphoton.2013.177>. URL <https://doi.org/10.1038/nphoton.2013.177>.
- Francesco Albarelli and Rafał Demkowicz-Dobrzański. Probe incompatibility in multiparameter noisy quantum metrology. *Phys. Rev. X*, 12:011039, Mar 2022. doi: 10.1103/PhysRevX.12.011039. URL <https://link.aps.org/doi/10.1103/PhysRevX.12.011039>.
- Francesco Albarelli and Rafał Demkowicz-Dobrzański. Probe incompatibility in multiparameter noisy quantum channel estimation. *arXiv preprint arXiv:2104.11264*, 2021. URL <https://arxiv.org/abs/2104.11264>.
- S. Alipour and A. T. Rezakhani. Extended convexity of quantum fisher information in quantum metrology. *Phys. Rev. A*, 91:042104, Apr 2015. doi: 10.1103/PhysRevA.91.042104. URL <https://link.aps.org/doi/10.1103/PhysRevA.91.042104>.
- Petr M. Anisimov, Gretchen M. Raterman, Aravind Chiruvelli, William N. Plick, Sean D. Huver, Hwang Lee, and Jonathan P. Dowling. Quantum metrology with two-mode squeezed vacuum: Parity detection beats the heisenberg limit. *Phys. Rev. Lett.*, 104:103602, Mar 2010. doi: 10.1103/PhysRevLett.104.103602. URL <https://link.aps.org/doi/10.1103/PhysRevLett.104.103602>.
- E. Bagan, M. Baig, A. Brey, R. Muñoz Tapia, and R. Tarrach. Optimal strategies for sending information through a quantum channel. *Phys. Rev. Lett.*, 85:5230–5233, Dec 2000. doi: 10.1103/PhysRevLett.85.5230. URL <https://link.aps.org/doi/10.1103/PhysRevLett.85.5230>.
- E. Bagan, M. Baig, A. Brey, R. Muñoz Tapia, and R. Tarrach. Optimal encoding and decoding of a spin direction. *Phys. Rev. A*, 63:052309, Apr 2001. doi: 10.1103/PhysRevA.63.052309. URL <https://link.aps.org/doi/10.1103/PhysRevA.63.052309>.
- E. Bagan, M. Baig, and R. Muñoz Tapia. Quantum reverse engineering and reference-frame alignment without nonlocal correlations. *Phys. Rev. A*, 70:030301, Sep 2004. doi: 10.1103/PhysRevA.70.030301. URL <https://link.aps.org/doi/10.1103/PhysRevA.70.030301>.
- A Bandilla, H Paul, and H H Ritze. Realistic quantum states of light with minimum phase uncertainty. *Quantum Optics: Journal of the European Optical Society Part B*, 3(5):267–282, oct 1991. doi: 10.1088/0954-8998/3/5/003. URL <https://doi.org/10.1088/0954-8998/3/5/003>.
- Tillmann Baumgratz and Animesh Datta. Quantum enhanced estimation of a multidimensional field. *Phys. Rev. Lett.*, 116:030801, Jan 2016. doi: 10.1103/PhysRevLett.116.030801. URL <https://link.aps.org/doi/10.1103/PhysRevLett.116.030801>.
- José Beltrán and Alfredo Luis. Breaking the heisenberg limit with inefficient detectors. *Phys. Rev.*

- A, 72:045801, Oct 2005. doi: 10.1103/PhysRevA.72.045801. URL <https://link.aps.org/doi/10.1103/PhysRevA.72.045801>.
- D. W. Berry and H. M. Wiseman. Optimal states and almost optimal adaptive measurements for quantum interferometry. *Phys. Rev. Lett.*, 85:5098–5101, Dec 2000. doi: 10.1103/PhysRevLett.85.5098. URL <https://link.aps.org/doi/10.1103/PhysRevLett.85.5098>.
- D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman. How to perform the most accurate possible phase measurements. *Phys. Rev. A*, 80:052114, Nov 2009. doi: 10.1103/PhysRevA.80.052114. URL <https://link.aps.org/doi/10.1103/PhysRevA.80.052114>.
- Dominic W. Berry, Michael J. W. Hall, and Howard M. Wiseman. Stochastic heisenberg limit: Optimal estimation of a fluctuating phase. *Phys. Rev. Lett.*, 111:113601, Sep 2013. URL <https://link.aps.org/doi/10.1103/PhysRevLett.111.113601>.
- Iwo Białynicki-Birula and Jerzy Mycielski. Uncertainty relations for information entropy in wave mechanics. *Communications in Mathematical Physics*, 44(2):129–132, 1975. doi: <https://doi.org/10.1007/BF01608825>. URL <https://doi.org/10.1007/BF01608825>.
- N. A. Bogomolov. Minimax measurements in a general statistical decision theory. *Theory of Probability & Its Applications*, 26(4):787–795, 1982. doi: 10.1137/1126084. URL <https://doi.org/10.1137/1126084>.
- Sergio Boixo, Steven T. Flammia, Carlton M. Caves, and JM Geremia. Generalized limits for single-parameter quantum estimation. *Phys. Rev. Lett.*, 98:090401, Feb 2007. doi: 10.1103/PhysRevLett.98.090401. URL <https://link.aps.org/doi/10.1103/PhysRevLett.98.090401>.
- Samuel L. Braunstein and Carlton M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72:3439–3443, May 1994. doi: 10.1103/PhysRevLett.72.3439. URL <https://link.aps.org/doi/10.1103/PhysRevLett.72.3439>.
- V. Bužek, R. Derka, and S. Massar. Optimal quantum clocks. *Phys. Rev. Lett.*, 82:2207–2210, Mar 1999. doi: 10.1103/PhysRevLett.82.2207. URL <https://link.aps.org/doi/10.1103/PhysRevLett.82.2207>.
- Carlton M Caves. Quantum-mechanical noise in an interferometer. *Phys. Rev. D*, 23(8):1693, 1981. URL <https://journals.aps.org/prd/pdf/10.1103/PhysRevD.23.1693>.
- Krzysztof Chabuda, Jacek Dziarmaga, Tobias J Osborne, and Rafał Demkowicz-Dobrzański. Tensor-network approach for quantum metrology in many-body quantum systems. *Nature communications*, 11(1):1–12, 2020. doi: <https://doi.org/10.1038/s41467-019-13735-9>. URL <https://doi.org/10.1038/s41467-019-13735-9>.
- Hongzhen Chen and Haidong Yuan. Optimal joint estimation of multiple rabi frequencies. *Phys. Rev. A*, 99:032122, Mar 2019. doi: 10.1103/PhysRevA.99.032122. URL <https://link.aps.org/doi/10.1103/PhysRevA.99.032122>.

10.1103/PhysRevA.99.032122.

- G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi. Efficient use of quantum resources for the transmission of a reference frame. *Phys. Rev. Lett.*, 93:180503, Oct 2004. doi: 10.1103/PhysRevLett.93.180503. URL <https://link.aps.org/doi/10.1103/PhysRevLett.93.180503>.
- G. Chiribella, G. M. D'Ariano, and M. F. Sacchi. Optimal estimation of group transformations using entanglement. *Phys. Rev. A*, 72:042338, Oct 2005. doi: 10.1103/PhysRevA.72.042338. URL <https://link.aps.org/doi/10.1103/PhysRevA.72.042338>.
- G. Chiribella, G. M. D'Ariano, and P. Perinotti. Quantum circuit architecture. *Phys. Rev. Lett.*, 101:060401, Aug 2008a. doi: 10.1103/PhysRevLett.101.060401. URL <https://link.aps.org/doi/10.1103/PhysRevLett.101.060401>.
- Giulio Chiribella, Giacomo M. D'Ariano, and Paolo Perinotti. Memory effects in quantum channel discrimination. *Phys. Rev. Lett.*, 101:180501, Oct 2008b. doi: 10.1103/PhysRevLett.101.180501. URL <https://link.aps.org/doi/10.1103/PhysRevLett.101.180501>.
- Dariusz Chruściński and Andrzej Kossakowski. Spectral conditions for positive maps. *Communications in Mathematical Physics*, 290(3):1051–1064, 2009. URL <https://doi.org/10.1007/s00220-009-0790-8>.
- Joel E. Cohen, Shmuel Friedland, Tosio Kato, and Frank P. Kelly. Eigenvalue inequalities for products of matrix exponentials. *Linear Algebra and its Applications*, 45:55–95, 1982. ISSN 0024-3795. doi: [https://doi.org/10.1016/0024-3795\(82\)90211-7](https://doi.org/10.1016/0024-3795(82)90211-7). URL <https://www.sciencedirect.com/science/article/pii/0024379582902117>.
- Jack Davis, Meenu Kumari, Robert B. Mann, and Shohini Ghose. Wigner negativity in spin- j systems. *Phys. Rev. Research*, 3:033134, Aug 2021. doi: 10.1103/PhysRevResearch.3.033134. URL <https://link.aps.org/doi/10.1103/PhysRevResearch.3.033134>.
- Christian L Degen, F Reinhard, and P Cappellaro. Quantum sensing. *Rev. Mod. Phys.*, 89(3):035002, 2017. URL <https://journals.aps.org/rmp/pdf/10.1103/RevModPhys.89.035002>.
- R. Demkowicz-Dobrzanski, M. Jarzyna, and J. Kołodyński. Quantum limits in optical interferometry. In Emil Wolf, editor, *Prog. Optics*, volume 60, pages 345–435. Elsevier, 2015. doi: 10.1016/bs.po.2015.02.003. URL <https://www.sciencedirect.com/science/article/pii/S0079663815000049>.
- Rafał Demkowicz-Dobrzański and Lorenzo Maccone. Using entanglement against noise in quantum metrology. *Phys. Rev. Lett.*, 113(25):250801, 2014. URL <https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.113.250801>.
- Rafał Demkowicz-Dobrzański, Jan Kołodyński, and Mădălin Guță. The elusive heisenberg limit in quantum-enhanced metrology. *Nat. Commun.*, 3:1063, 2012. URL <https://www.nature.com/articles/ncomms2067>.

- Rafał Demkowicz-Dobrzański, Wojciech Górecki, and Mădălin Guță. Multi-parameter estimation beyond quantum fisher information. *Journal of Physics A: Mathematical and Theoretical*, 53(36): 363001, 2020. URL <https://doi.org/10.1088/1751-8121/ab8ef3>.
- Rafał Demkowicz-Dobrzański, Marcin Jarzyna, and Jan Kołodyński. Chapter four - quantum limits in optical interferometry. volume 60 of *Progress in Optics*, pages 345–435. Elsevier, 2015. doi: <https://doi.org/10.1016/bs.po.2015.02.003>. URL <https://www.sciencedirect.com/science/article/pii/S0079663815000049>.
- Akio Fujiwara and Hiroshi Imai. A fibre bundle over manifolds of quantum channels and its application to quantum statistics. *J. Phys. A: Math. Theor.*, 41(25):255304, 2008. URL <http://iopscience.iop.org/article/10.1088/1751-8113/41/25/255304/pdf>.
- Wenchao Ge, Kurt Jacobs, Zachary Eldredge, Alexey V Gorshkov, and Michael Foss-Feig. Distributed quantum metrology with linear networks and separable inputs. *Phys. Rev. Lett.*, 121(4): 043604, 2018. URL <https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.121.043604>.
- Marco G. Genoni, Stefano Mancini, and Alessio Serafini. Optimal feedback control of linear quantum systems in the presence of thermal noise. *Phys. Rev. A*, 87:042333, Apr 2013. doi: 10.1103/PhysRevA.87.042333. URL <https://link.aps.org/doi/10.1103/PhysRevA.87.042333>.
- Manuel Gessner, Luca Pezzè, and Augusto Smerzi. Sensitivity bounds for multiparameter quantum metrology. *Phys. Rev. Lett.*, 121:130503, Sep 2018. doi: 10.1103/PhysRevLett.121.130503. URL <https://link.aps.org/doi/10.1103/PhysRevLett.121.130503>.
- RD Gill and S Massar. State estimation for large ensembles. *Phys. Rev. A*, 61:042312, 2000. URL <https://journals.aps.org/prl/pdf/10.1103/PhysRevA.61.042312>.
- Richard D Gill. Conciliation of bayes and pointwise quantum state estimation. In *Quantum Stochastics and Information: Statistics, Filtering and Control*, pages 239–261. World Scientific, 2008.
- Richard D. Gill and Boris Y. Levit. Applications of the van trees inequality: A bayesian cramér-rao bound. *Bernoulli*, 1(1/2):59–79, 1995. ISSN 13507265. URL <http://www.jstor.org/stable/3318681>.
- Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum limits to dynamical evolution. *Phys. Rev. A*, 67:052109, May 2003. doi: 10.1103/PhysRevA.67.052109. URL <https://link.aps.org/doi/10.1103/PhysRevA.67.052109>.
- Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Phys. Rev. Lett.*, 96(1):010401, 2006. URL <https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.96.010401>.
- Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Advances in quantum metrology. *Nat. Photonics*, 5(4):222–229, 2011. URL <https://www.nature.com/articles/nphoton.2011.35>.

- Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum measurement bounds beyond the uncertainty relations. *Phys. Rev. Lett.*, 108:260405, Jun 2012. doi: 10.1103/PhysRevLett.108.260405. URL <https://link.aps.org/doi/10.1103/PhysRevLett.108.260405>.
- Aaron Z. Goldberg, Ilaria Gianani, Marco Barbieri, Fabio Sciarrino, Aephraim M. Steinberg, and Nicolò Spagnolo. Multiphase estimation without a reference mode. *Phys. Rev. A*, 102:022230, Aug 2020. doi: 10.1103/PhysRevA.102.022230. URL <https://link.aps.org/doi/10.1103/PhysRevA.102.022230>.
- Wojciech Górecki and Rafał Demkowicz-Dobrzański. Multiple-phase quantum interferometry: Real and apparent gains of measuring all the phases simultaneously. *Phys. Rev. Lett.*, 128:040504, Jan 2022a. doi: 10.1103/PhysRevLett.128.040504. URL <https://link.aps.org/doi/10.1103/PhysRevLett.128.040504>.
- Wojciech Górecki and Rafał Demkowicz-Dobrzański. Multiparameter quantum metrology in the heisenberg limit regime: Many-repetition scenario versus full optimization. *Phys. Rev. A*, 106:012424, Jul 2022b. doi: 10.1103/PhysRevA.106.012424. URL <https://link.aps.org/doi/10.1103/PhysRevA.106.012424>.
- Wojciech Górecki, Rafał Demkowicz-Dobrzański, Howard M. Wiseman, and Dominic W. Berry. π -corrected heisenberg limit. *Phys. Rev. Lett.*, 124:030501, Jan 2020. doi: 10.1103/PhysRevLett.124.030501. URL <https://link.aps.org/doi/10.1103/PhysRevLett.124.030501>.
- Wojciech Górecki, Sisi Zhou, Liang Jiang, and Rafał Demkowicz-Dobrzański. Optimal probes and error-correction schemes in multi-parameter quantum metrology. *Quantum*, 4:288, July 2020. ISSN 2521-327X. doi: 10.22331/q-2020-07-02-288. URL <https://doi.org/10.22331/q-2020-07-02-288>.
- Wojciech Górecki, Alberto Riccardi, and Lorenzo Maccone. Quantum metrology of noisy spreading channels. *Phys. Rev. Lett.*, 129:240503, Dec 2022. doi: 10.1103/PhysRevLett.129.240503. URL <https://link.aps.org/doi/10.1103/PhysRevLett.129.240503>.
- Mădălin Guță and Jonas Kahn. Local asymptotic normality for qubit states. *Phys. Rev. A*, 73:052108, May 2006. doi: 10.1103/PhysRevA.73.052108. URL <https://link.aps.org/doi/10.1103/PhysRevA.73.052108>.
- Gregory Gundersen. Asymptotic normality of maximum likelihood estimators. 2019. URL <https://gregorygundersen.com/blog/2019/11/28/asymptotic-normality-mle/>.
- Mădălin Guță, Bas Janssens, and Jonas Kahn. Optimal estimation of qubit states with continuous time measurements. *Communications in Mathematical Physics*, 277(1):127–160, 2008. doi: <https://doi.org/10.1007/s00220-007-0357-5>. URL <https://doi.org/10.1007/s00220-007-0357-5>.
- Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641,

2017. doi: 10.1109/TIT.2017.2719044. URL <https://doi.org/10.1109/TIT.2017.2719044>.
- Jaroslav Hájek. Local asymptotic minimax and admissibility in estimation. In *Theory of Statistics*, pages 175–194. University of California Press, 1972.
- Michael J. W. Hall and Howard M. Wiseman. Does nonlinear metrology offer improved resolution? answers from quantum information theory. *Phys. Rev. X*, 2:041006, Oct 2012. doi: 10.1103/PhysRevX.2.041006. URL <https://link.aps.org/doi/10.1103/PhysRevX.2.041006>.
- Michael J. W. Hall, Dominic W. Berry, Marcin Zwierz, and Howard M. Wiseman. Universality of the heisenberg limit for estimates of random phase shifts. *Phys. Rev. A*, 85:041802, Apr 2012. doi: 10.1103/PhysRevA.85.041802. URL <https://link.aps.org/doi/10.1103/PhysRevA.85.041802>.
- Masahito Hayashi. Parallel treatment of estimation of $\text{su}(2)$ and phase estimation. *Physics Letters A*, 354(3):183–189, 2006. ISSN 0375-9601. doi: <https://doi.org/10.1016/j.physleta.2006.01.043>. URL <https://www.sciencedirect.com/science/article/pii/S0375960106001332>.
- Masahito Hayashi. Comparison between the cramer-rao and the mini-max approaches in quantum channel estimation. *Communications in mathematical physics*, 304(3):689–709, 2011. doi: <https://doi.org/10.1007/s00220-011-1239-4>. URL <https://doi.org/10.1007/s00220-011-1239-4>.
- Masahito Hayashi. Fourier analytic approach to quantum estimation of group action. *Communications in Mathematical Physics*, 347(1):3–82, 2016. doi: <https://doi.org/10.1007/s00220-016-2738-0>. URL <https://doi.org/10.1007/s00220-016-2738-0>.
- W Heitler. The quantum theory of radiation, 3rd edn, clarendon. 1954.
- BL Higgins, DW Berry, SD Bartlett, MW Mitchell, HM Wiseman, and GJ Pryde. Demonstrating heisenberg-limited unambiguous phase estimation without adaptive measurements. *New Journal of Physics*, 11(7):073023, 2009. doi: 10.1088/1367-2630/11/7/073023. URL <https://doi.org/10.1088/1367-2630/11/7/073023>.
- Brendon L Higgins, Dominic W Berry, Stephen D Bartlett, Howard M Wiseman, and Geoff J Pryde. Entanglement-free heisenberg-limited phase estimation. *Nature*, 450(7168):393–396, 2007. doi: <https://doi.org/10.1038/nature06257>. URL <https://doi.org/10.1038/nature06257>.
- Le Bin Ho, Hideaki Hakoshima, Yuichiro Matsuzaki, Masayuki Matsuzaki, and Yasushi Kondo. Multiparameter quantum estimation under dephasing noise. *Phys. Rev. A*, 102:022602, Aug 2020. doi: 10.1103/PhysRevA.102.022602. URL <https://link.aps.org/doi/10.1103/PhysRevA.102.022602>.
- A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory*. North Holland, Amsterdam, 1982. doi: 10.1007/978-88-7642-378-9. URL <https://doi.org/10.1007/978-88-7642-378-9>.
- M. J. Holland and K. Burnett. Interferometric detection of optical phase shifts at the heisenberg

- limit. *Phys. Rev. Lett.*, 71:1355–1358, Aug 1993. doi: 10.1103/PhysRevLett.71.1355. URL <https://link.aps.org/doi/10.1103/PhysRevLett.71.1355>.
- Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Reviews of modern physics*, 81(2):865, 2009. doi: <https://doi.org/10.1103/RevModPhys.81.865>. URL <https://doi.org/10.1103/RevModPhys.81.865>.
- Peter C. Humphreys, Marco Barbieri, Animesh Datta, and Ian A. Walmsley. Quantum enhanced multiple phase estimation. *Phys. Rev. Lett.*, 111:070403, Aug 2013a. doi: 10.1103/PhysRevLett.111.070403. URL <https://link.aps.org/doi/10.1103/PhysRevLett.111.070403>.
- Peter C Humphreys, Marco Barbieri, Animesh Datta, and Ian A Walmsley. Quantum enhanced multiple phase estimation. *Phys. Rev. Lett.*, 111(7):070403, 2013b. URL <https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.111.070403>.
- Hiroshi Imai and Masahito Hayashi. Fourier analytic approach to phase estimation in quantum systems. *New Journal of Physics*, 11(4):043034, apr 2009. doi: 10.1088/1367-2630/11/4/043034. URL <https://doi.org/10.1088/1367-2630/11/4/043034>.
- Yumang Jing, Matteo Fadel, Valentin Ivannikov, and Tim Byrnes. Split spin-squeezed bose–einstein condensates. *New Journal of Physics*, 21(9):093038, sep 2019. doi: 10.1088/1367-2630/ab3fcf. URL <https://dx.doi.org/10.1088/1367-2630/ab3fcf>.
- Tomasz Kaftal and Rafał Demkowicz-Dobrzański. Usefulness of an enhanced kitaev phase-estimation algorithm in quantum metrology and computation. *Phys. Rev. A*, 90:062313, Dec 2014. doi: 10.1103/PhysRevA.90.062313. URL <https://link.aps.org/doi/10.1103/PhysRevA.90.062313>.
- Jonas Kahn. Fast rate estimation of a unitary operation in $SU(d)$. *Phys. Rev. A*, 75:022326, Feb 2007. doi: 10.1103/PhysRevA.75.022326. URL <https://link.aps.org/doi/10.1103/PhysRevA.75.022326>.
- J. Kaiser and R. Schafer. On the use of the $i0$ -sinh window for spectrum analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(1):105–107, 1980. doi: 10.1109/TASSP.1980.1163349. URL <https://ieeexplore.ieee.org/document/1163349>.
- Piotr Kolenderski and Rafal Demkowicz-Dobrzanski. Optimal state for keeping reference frames aligned and the platonic solids. *Phys. Rev. A*, 78:052333, Nov 2008. doi: 10.1103/PhysRevA.78.052333. URL <https://link.aps.org/doi/10.1103/PhysRevA.78.052333>.
- B. Kraus. Local unitary equivalence of multipartite pure states. *Phys. Rev. Lett.*, 104:020504, Jan 2010. doi: 10.1103/PhysRevLett.104.020504. URL <https://link.aps.org/doi/10.1103/PhysRevLett.104.020504>.
- H R Krishnamurthy, H S Mani, and H C Verma. Exact solution of the schrodinger equation for a particle in a tetrahedral box. *Journal of Physics A: Mathematical and General*, 15(7):2131–2137, jul 1982. doi: 10.1088/0305-4470/15/7/024. URL <https://doi.org/10.1088/0305-4470/15/7/024>.

- Naoto Kura and Masahito Ueda. Finite-error metrological bounds on multiparameter hamiltonian estimation. *Phys. Rev. A*, 97:012101, Jan 2018. doi: 10.1103/PhysRevA.97.012101. URL <https://link.aps.org/doi/10.1103/PhysRevA.97.012101>.
- Stanislaw Kurdzialek, Wojciech Gorecki, Francesco Albarelli, and Rafal Demkowicz-Dobrzanski. Using adaptiveness and causal superpositions against noise in quantum metrology. *arXiv preprint arXiv:2212.08106*, 2022.
- Wai-Kee Li. A particle in an isocles right triangle. *Journal of Chemical Education*, 61(12):1034, 1984. URL <https://doi.org/10.1021/ed061p1034>.
- The LIGO Scientific Collaboration LIGO. A gravitational wave observatory operating beyond the quantum shot-noise limit. *Nature Physics*, 7(12):962–965, 2011. doi: <https://doi.org/10.1038/nphys2083>. URL <https://doi.org/10.1038/nphys2083>.
- Jing Liu and Haidong Yuan. Control-enhanced multiparameter quantum estimation. *Phys. Rev. A*, 96:042114, Oct 2017. doi: 10.1103/PhysRevA.96.042114. URL <https://link.aps.org/doi/10.1103/PhysRevA.96.042114>.
- Qiushi Liu, Zihao Hu, Haidong Yuan, and Yuxiang Yang. Strict hierarchy of strategies for non-asymptotic quantum metrology. *arXiv preprint arXiv:2203.09758*, 2022.
- A. Luis and J. Peřina. Optimum phase-shift estimation and the quantum description of the phase difference. *Phys. Rev. A*, 54:4564–4570, Nov 1996. doi: 10.1103/PhysRevA.54.4564. URL <https://link.aps.org/doi/10.1103/PhysRevA.54.4564>.
- Katarzyna Macieszczak, Martin Fraas, and Rafał Demkowicz-Dobrzański. Bayesian quantum frequency estimation in presence of collective dephasing. *New Journal of Physics*, 16(11):113002, oct 2014. doi: 10.1088/1367-2630/16/11/113002. URL <https://dx.doi.org/10.1088/1367-2630/16/11/113002>.
- L. Mandelstam and Ig. Tamm. *The Uncertainty Relation Between Energy and Time in Non-relativistic Quantum Mechanics*, pages 115–123. Springer Berlin Heidelberg, Berlin, Heidelberg, 1991. ISBN 978-3-642-74626-0. doi: 10.1007/978-3-642-74626-0_8. URL https://doi.org/10.1007/978-3-642-74626-0_8.
- Norman Margolus and Lev B. Levitin. The maximum speed of dynamical evolution. *Physica D: Nonlinear Phenomena*, 120(1):188–195, 1998. ISSN 0167-2789. doi: [https://doi.org/10.1016/S0167-2789\(98\)00054-2](https://doi.org/10.1016/S0167-2789(98)00054-2). URL <https://www.sciencedirect.com/science/article/pii/S0167278998000542>. Proceedings of the Fourth Workshop on Physics and Consumption.
- Keiji Matsumoto. A new approach to the cramér-rao-type bound of the pure-state model. *J. Phys. A.: Math. Theor.*, 35(13):3111, 2002. URL <http://iopscience.iop.org/article/10.1088/0305-4470/35/13/307/meta>.

- Morgan W Mitchell, Jeff S Lundeen, and Aephraem M Steinberg. Super-resolving phase measurements with a multiphoton entangled state. *Nature*, 429(6988):161–164, 2004. doi: <https://doi.org/10.1038/nature02493>. URL <https://doi.org/10.1038/nature02493>.
- H Nagaoka and M Hayashi. *Asymptotic Theory Of Quantum Statistical Inference: Selected Papers*. World Scientific Singapore, 2005. doi: <https://doi.org/10.1142/5630>. URL <https://doi.org/10.1142/5630>.
- Whitney K Newey and Daniel McFadden. Chapter 36 large sample estimation and hypothesis testing. volume 4 of handbook of econometrics. *Elsevier*, 12:2111–2245, 1994. doi: [https://doi.org/10.1016/S1573-4412\(05\)80005-4](https://doi.org/10.1016/S1573-4412(05)80005-4). URL [https://doi.org/10.1016/S1573-4412\(05\)80005-4](https://doi.org/10.1016/S1573-4412(05)80005-4).
- Rosanna Nichols, Pietro Liuzzo-Scorpo, Paul A. Knott, and Gerardo Adesso. Multiparameter gaussian quantum metrology. *Phys. Rev. A*, 98:012114, Jul 2018. doi: 10.1103/PhysRevA.98.012114. URL <https://link.aps.org/doi/10.1103/PhysRevA.98.012114>.
- Artur Niezgodą, Dariusz Kajtoch, Joanna Dziekańska, and Emilia Witkowska. Optimal quantum interferometry robust to detection noise using spin-1 atomic condensates. *New Journal of Physics*, 21(9):093037, 2019. doi: 10.1088/1367-2630/ab4099. URL <https://doi.org/10.1088/1367-2630/ab4099>.
- Mattteo G. A. Paris. Quantum estimation for quantum technologies. *Int. J. Quantum Inf.*, 07:125–137, 2009. doi: 10.1142/S0219749909004839. URL <https://doi.org/10.1142/S0219749909004839>.
- Luca Pezzé. Sub-heisenberg phase uncertainties. *Phys. Rev. A*, 88:060101, Dec 2013. doi: 10.1103/PhysRevA.88.060101. URL <https://link.aps.org/doi/10.1103/PhysRevA.88.060101>.
- Luca Pezzè, Augusto Smerzi, Markus K. Oberthaler, Roman Schmied, and Philipp Treutlein. Quantum metrology with nonclassical states of atomic ensembles. *Rev. Mod. Phys.*, 90:035005, Sep 2018. doi: 10.1103/RevModPhys.90.035005. URL <https://link.aps.org/doi/10.1103/RevModPhys.90.035005>.
- S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd. Advances in photonic quantum sensing. *Nat. Photonics*, 12(12):724–733, 2018. ISSN 1749-4893. doi: 10.1038/s41566-018-0301-6. URL <https://doi.org/10.1038/s41566-018-0301-6>.
- Jr. R. P. Boas and M. Kac. Inequalities for Fourier transforms of positive functions. *Duke Mathematical Journal*, 12(1):189 – 206, 1945. doi: 10.1215/S0012-7094-45-01215-4. URL <https://doi.org/10.1215/S0012-7094-45-01215-4>.
- Sammy Ragy, Marcin Jarzyna, and Rafał Demkowicz-Dobrzański. Compatibility in multiparameter quantum metrology. *Phys. Rev. A*, 94:052108, Nov 2016. doi: 10.1103/PhysRevA.94.052108. URL <https://link.aps.org/doi/10.1103/PhysRevA.94.052108>.
- Angel Rivas and Alfredo Luis. Sub-heisenberg estimation of non-random phase shifts. *New Journal*

- of *Physics*, 14(9):093052, 2012. URL <https://link.aps.org/doi/10.1088/1367-2630/14/9/093052>.
- H. P. Robertson. An indeterminacy relation for several observables and its classical interpretation. *Phys. Rev.*, 46:794–801, Nov 1934. doi: 10.1103/PhysRev.46.794. URL <https://link.aps.org/doi/10.1103/PhysRev.46.794>.
- R. Schnabel. Squeezed states of light and their applications in laser interferometers. *Phys. Rep.*, 684:1 – 51, 2017. ISSN 0370-1573. URL <http://www.sciencedirect.com/science/article/pii/S0370157317300595>.
- Erwin Schrödinger. *Zum heisenbergschen unschärfeprinzip*. Akademie der Wissenschaften, 1930.
- G.S. Summy and D.T. Pegg. Phase optimized quantum states of light. *Optics Communications*, 77(1):75–79, 1990. ISSN 0030-4018. doi: [https://doi.org/10.1016/0030-4018\(90\)90464-5](https://doi.org/10.1016/0030-4018(90)90464-5). URL <https://www.sciencedirect.com/science/article/pii/0030401890904645>.
- Jun Suzuki. Nuisance parameter problem in quantum estimation theory: Tradeoff relation and qubit examples. *Journal of Physics A: Mathematical and Theoretical*, 53(26):264001, 2020. doi: 10.1088/1751-8121/ab8672. URL <https://doi.org/10.1088/1751-8121/ab8672>.
- Jun Suzuki, Yuxiang Yang, and Masahito Hayashi. Quantum state estimation with nuisance parameters. *Journal of Physics A: Mathematical and Theoretical*, 53(45):453001, 2020. URL <https://doi.org/10.1088/1751-8121/ab8b78>.
- Geza Toth and Iagoba Apellaniz. Quantum metrology from a quantum information science perspective. *J. Phys. A: Math. Theor.*, 47(42):424006, 2014. URL <http://stacks.iop.org/1751-8121/47/i=42/a=424006>.
- Mankei Tsang, Howard M. Wiseman, and Carlton M. Caves. Fundamental quantum limit to waveform estimation. *Phys. Rev. Lett.*, 106:090401, Mar 2011. URL <http://link.aps.org/doi/10.1103/PhysRevLett.106.090401>.
- Mankei Tsang, Ranjith Nair, and Xiao-Ming Lu. Quantum theory of superresolution for two incoherent optical point sources. *Phys. Rev. X*, 6(3):031033, 2016. URL <https://journals.aps.org/prx/pdf/10.1103/PhysRevX.6.031033>.
- J W Turner. On the quantum particle in a polyhedral box. *Journal of Physics A: Mathematical and General*, 17(14):2791–2797, oct 1984. doi: 10.1088/0305-4470/17/14/022. URL <https://doi.org/10.1088/0305-4470/17/14/022>.
- A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256. URL <https://doi.org/10.1017/CBO9780511802256>.
- Fabian Wolf, Chunyan Shi, Jan C Heip, Manuel Gessner, Luca Pezzè, Augusto Smerzi, Mar-

- ius Schulte, Klemens Hammerer, and Piet O Schmidt. Motional fock states for quantum-enhanced amplitude and phase measurements with trapped ions. *Nature communications*, 10(1):1–8, 2019. doi: <https://doi.org/10.1038/s41467-019-10576-4>. URL <https://doi.org/10.1038/s41467-019-10576-4>.
- Yuxiang Yang, Giulio Chiribella, and Masahito Hayashi. Attaining the ultimate precision limit in quantum state estimation. *Communications in Mathematical Physics*, 368(1):223–293, 2019. doi: 10.1007/s00220-019-03433-4. URL <https://doi.org/10.1007/s00220-019-03433-4>.
- Ying-Wang, Li-Yun Hu, Zhi-Ming Zhang, and Chao-Ping Wei. Simultaneously estimating two phases in three-mode nonlinear interferometer. *Optics Communications*, 427:250–256, 2018. ISSN 0030-4018. doi: <https://doi.org/10.1016/j.optcom.2018.06.072>. URL <https://www.sciencedirect.com/science/article/pii/S0030401818305637>.
- Rozhin Yousefjani, Rosanna Nichols, Shahriar Salimi, and Gerardo Adesso. Estimating phase with a random generator: Strategies and resources in multiparameter quantum metrology. *Phys. Rev. A*, 95:062307, Jun 2017. doi: 10.1103/PhysRevA.95.062307. URL <https://link.aps.org/doi/10.1103/PhysRevA.95.062307>.
- Haidong Yuan. Sequential feedback scheme outperforms the parallel scheme for hamiltonian parameter estimation. *Phys. Rev. Lett.*, 117:160801, Oct 2016. doi: 10.1103/PhysRevLett.117.160801. URL <https://link.aps.org/doi/10.1103/PhysRevLett.117.160801>.
- Xiaobin Zhao, Yuxiang Yang, and Giulio Chiribella. Quantum metrology with indefinite causal order. *Phys. Rev. Lett.*, 124:190503, May 2020. doi: 10.1103/PhysRevLett.124.190503. URL <https://link.aps.org/doi/10.1103/PhysRevLett.124.190503>.
- Sisi Zhou and Liang Jiang. Asymptotic theory of quantum channel estimation. *PRX Quantum*, 2:010343, Mar 2021. doi: 10.1103/PRXQuantum.2.010343. URL <https://link.aps.org/doi/10.1103/PRXQuantum.2.010343>.
- Marcin Zwierz, Carlos A. Pérez-Delgado, and Pieter Kok. General optimality of the heisenberg limit for quantum metrology. *Phys. Rev. Lett.*, 105:180402, Oct 2010. doi: 10.1103/PhysRevLett.105.180402. URL <https://link.aps.org/doi/10.1103/PhysRevLett.105.180402>.
- Marcin Zwierz, Carlos A. Pérez-Delgado, and Pieter Kok. Erratum: General optimality of the heisenberg limit for quantum metrology [phys. rev. lett. 105, 180402 (2010)]. *Phys. Rev. Lett.*, 107:059904, Jul 2011. doi: 10.1103/PhysRevLett.107.059904. URL <https://link.aps.org/doi/10.1103/PhysRevLett.107.059904>.