Multiparameter quantum metrology in the Heisenberg limit regime: Many-repetition scenario versus full optimization

Wojciech Górecki and Rafał Demkowicz-Dobrzański
Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warsaw, Poland

(Received 14 March 2022; accepted 24 June 2022; published 18 July 2022)

We discuss the Heisenberg limit in the multiparameter metrology within two different paradigms: one where the measurement is repeated many times (so the Cramér-Rao bound is guaranteed to be asymptotically saturable). and the second one where all the resources are allocated into one experimental realization (analyzed with the mimimax approach). We investigate the potential advantage of measuring all the parameters simultaneously compared to estimating them individually, while spending the same total amount of resources. We show that in general the existence of such an advantage, its magnitude, and conditions under which it occurs depends on which of the two paradigms has been chosen. In particular, for the problem of magnetic field sensing using N entangled spin $\frac{1}{2}$, we show that the predictions based purely on the Cramér-Rao formalism may be overly pessimistic in this matter: the minimax approach reveals the superiority of measuring all the parameters jointly whereas the Cramér-Rao approach indicates lack of such an advantage.

DOI: 10.1103/PhysRevA.106.012424

I. INTRODUCTION

Quantum mechanics opens up new possibilities in metrology, enabling the use of coherence and entanglement to increase measurement precision [1–9]. The most prominent example of this is the ability to overcome the shot-noise-limit linear scaling of the estimation precision with the number of resources n used in measurement (which can be understood as number of photons, total energy, total time, etc.) and obtain a quadratic scaling, the so-called Heisenberg scaling [10–19]. Even if presence of decoherence makes the Heisenberg scaling fragile and virtually impossible to preserve in the asymptotic limit [20,21], for many models the noise may be completely or partially canceled by applying proper quantum error correction protocols [22–26], which allows for the observation of the quadratic precision scaling in certain finite-resource regimes.

If an experiment aimed at estimation a parameter θ is repeated k times and involves the use of n resources in each repetition, then provided the Heisenberg scaling holds, the variance of the estimator will scale as

$$\Delta^2\tilde{\theta} \propto \frac{1}{kn^2}. \tag{1}$$

If k is sufficiently large, then the problem may be successfully analyzed with the use of the concept of quantum Fisher information (QFI) and the related Cramér-Rao (CR) bound (which is proven to be tight in the limit $k \to \infty$). However, as pointed out in [27–32], a subtle problem appears if one wants to discuss the best precision achievable when all the available resources $N = nk$ are used optimally, which we will refer to as the actual Heisenberg limit.

When inspecting Eq. (1) it is apparent that, when $N = nk$ is kept fixed, one should accumulate as much resources as possible in a single repetition of an experiment and therefore increase n at the expense of smaller k. Unfortunately, in general it is not clear what is the minimal number of repetition k needed to saturate the CR in practice (which may be different for various models), and hence the sole notion of the QFI does not provide a full understanding of the problem.

This case was broadly discussed for the problem of estimating phase shift in the interferometer using an N-photon state (within the Bayesian [33–35] or minimax [27] formalism) as well as for general problem of single-parameter unitary estimation [32]. It was shown that the optimal state is different than the one maximizing the QFI and that the final estimator's variance is π^2 times larger then the one resulting from the QFI-based analysis (see Sec. III A for more discussion).

This implies that whenever the Heisenberg scaling occurs, then in order to discuss the optimal measurement strategy, one needs to strictly define which paradigm is under consideration. The one where all resources $N = nk$ may be used in the optimal way and accumulated in a single experiment's realization (which in this work we analyze within minimax formalism, labeled by MM) or the second one, where the amount of resources used in single trial n is large but finite, and the whole experiment is repeated many times k (analyzed within Cramér-Rao formalism, labeled by CR). In the latter case, the limit $N \to \infty$ corresponds to $k \to \infty, n = \text{const}$. Only such a formulation allows us to apply the general argument about the asymptotical saturability of the CR bounds.

While the issues mentioned above appear now to be completely understood in a single-parameter estimation case, new questions and challenges arise when discussing multiparameter estimation models [36–42]. In some situations, a properly designed multiparameter estimation protocol allows to reduce the total error in estimation when compared with a strategy where all the parameters are measured separately in independently prepared experiments [43,44]. Heisenberg scaling in multiparameter metrology has been discussed in the literature using both paradigms. Many-repetition scenarios have been considered in [45–49] (using multiparameter quantum CR bound) and [26,50–52] (using tighter variants of quantum CR bound), while single-experiment scenario where the total amount of resources is limited has been analyzed mainly within the Bayesian paradigm for models with underlying group symmetry (covariant problems): SU(2)/U(1) [53,54], SU(2) [55–58], SO(3) [59], SU(d) [60]. The quantitative analysis of the relation between the results obtained within these two paradigms has started to be analyzed only very recently [61].

The most pressing question is whether a gain can be made by measuring all the parameters simultaneously instead of separately, while consuming the same total amount of resources? That the necessity of splitting these resources between experiments focusing on estimating a given parameter in the separate case will in general have different consequences in different paradigms [61]. If only the total amount of resources N is restricted, and p parameters are to be estimated, it is rather clear that one needs to spend $\sim N/p$ resources per parameters. Since we assume the quadratic scaling of precision with the amount of resource used, we may expect the scaling of the sum of variances to be

$$\sum_{i=1}^{p} \Delta^2\tilde{\theta}_i \propto p \times \frac{1}{(N/p)^2} = \frac{p^3}{N^2}. \tag{2}$$

In the many repetition scenario, different approaches to analyzing this problem may be found in modern literature. A quite common method is to compare the optimal cost obtainable with applying joint measurement with the one obtained by dividing amount of resources achievable in single trial n between all the parameters, with the total number of repetitions of the whole experiment k kept unchanged [45,46,62–64], which leads to $\sum_{i=1}^{p} \Delta^2\tilde{\theta}_i \propto p \times \frac{1}{k} \times \frac{1}{(n/p)^2} = \frac{p^3}{kn^2}$. However, one should notice that there is no point in dividing n, while we have $k \gg n$ trials at our disposal. It is much more efficient to use $\sim k/p$ trials for each parameter [65–67]. Only such formulation of the problem allows for a fair comparison between measuring parameters jointly vs separately, and guarantees that the eventual superiority of the first one comes directly from measuring parameters jointly, not from more efficient resources distribution. As a result we obtain

$$\sum_{i=1}^{p} \Delta^2\tilde{\theta}_i \propto p \times \frac{1}{k/p} \times \frac{1}{n^2} = \frac{p^2}{kn^2}, \tag{3}$$

which exhibits scaling with p with a different power than in Eq. (2).

The nontrivial question is as follows: If the sum of variances for the optimal joint measurement will follow a similar scaling, it may turn out that the existence of the advantage depends on paradigm chosen. As recently shown in [61], for the multiphase estimation problem in a multiarm interferometer, this issue does not lead to divergent conclusions, even if the scaling of the total cost with the number of the parameters depends on the paradigm, the potential advantage obtainable by measuring all of the parameters jointly instead of separately is very similar in both paradigms. In this paper we analyze this issue more generally and show that result is not universal; it may happen that existence or absence of the advantage indeed depends on the paradigm chosen.

The paper is organized as follows: In Sec. II we remind the basis of the quantum measurement theory, define precisely what we mean by "measuring parameters separately," and introduce mathematical formalism useful when analyzing the problem in both paradigms. In Sec. III we derive some general bounds for the achievable precision. Finally, in Sec. IV we study multiparameter estimation models representing a variety of magnetic field sensing tasks, check the tightness of introduced bounds, and discuss the relation between the results obtained within both paradigms. We also contrast the results of magnetic field sensing models with the multiple-arm interferometry case.

The examples are intended to present to the reader in a simple way the diversity and the complicity of the relationship between optimal results obtainable within separate and joint strategies in both discussed paradigms, therefore, most of them are easily solvable with basic algebra or are based on analyzing existing results [53–55,57]. However, in Sec. IV B we present also original results about optimal joint measurement within the minimax approach for spatially distributed magnetic field sensing (which is a specific example with commuting evolution generators). The end results are summarized in Table I and Sec. V.

II. PROBLEM FORMULATION

Let $\mathcal{E}_{\boldsymbol{\theta}}$ be a quantum channel depending on a vector of unknown parameters $\boldsymbol{\theta} = [\theta_1, \dots, \theta_p]^T$. The aim is to estimate the values of $\boldsymbol{\theta}$ by sending through the channel some initial state $\mathcal{E}_{\boldsymbol{\theta}}(\rho_{\text{in}}) = \rho_{\boldsymbol{\theta}}$, performing the measurement $\{M_x\}$ (satisfying $\int dx\, M_x = \mathbb{1}$) on the $\rho_{\boldsymbol{\theta}}$ [leading to probability distribution of the result x given by $p_{\boldsymbol{\theta}}(x) = \text{Tr}(M_x \rho_{\boldsymbol{\theta}})$] and assigning to outputs of the measurement proper values of estimators $\tilde{\theta}_i(x)$. For such a strategy we define the estimators covariance matrix by

$$\Sigma = \int dx\, p_{\boldsymbol{\theta}}(x)[\tilde{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}][\tilde{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}]^T, \tag{4}$$

and the aim is to minimize its trace, i.e., the sum of squared deviations of the estimator from the true value (for simplicity we will refer to them as variances, implicitly assuming that the estimators will likely be unbiased and their expectation value will coincide with the true value of the parameter) of all parameters:

$$\Delta^2\tilde{\boldsymbol{\theta}} := \sum_{i=1}^{p} \Delta^2\tilde{\theta}_i = \text{Tr}(\Sigma). \tag{5}$$

We will compare the value of the above variance at some reference point $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, keeping the condition that the measurement should work well also in some small neighborhood of this point (two alternative methods how to formalize this condition will be given later).

In the literature, a more general cost function is sometimes considered, where the covariance matrix is additionally

TABLE I. Optimal achievable sum of variances of estimated parameters for six different models with unitary evolution governed by the Hamiltonian $\sum \theta_i \Lambda_i$. Two paradigms are compared: the one where the channel $e^{i\sum \theta_i \Lambda_i}$ is used N times in the optimal way and the second where it is used n times in each single trial, which is repeated k times. All the values are presented in the limits of large k, n, N, and p (the exact formulas may be found in the main text). We analyze the optimal strategy when all p parameters are measured separately, as well as the case when they are measured jointly. The asymptotic values and inequalities are written for the parallel strategy case (if adaptiveness helps the minimal cost obtainable is written in a bracket). The constants in the two last columns are proven to be $0.63 \leqslant c_1 \leqslant 1$ and $1.89 \leqslant c_2 \leqslant 2$. The $\sigma_z^{(i)}$ in the definition of the Hamiltonian in the middle column is a shortcut for $\mathbb{1}^{\otimes i-1} \otimes \sigma_z/2 \otimes \mathbb{1}^{\otimes(p-i)}$. The relation between the costs obtained by different strategies in general depends on the paradigm chosen.

$\min \sum_{i=1}^{p} \Delta^2\tilde{\theta}_i$	model	magnetic field estimation					multiphase estimation	scaling
		1 component	2 components	3 components	1 component, spatial distribution			
					p-atoms layers, free spin orientation	single atoms, free position and spin		
	Hamiltonian	$\theta\sigma_z/2$	$\sum_{i\in\{y,z\}} \theta_i\sigma_i/2$	$\sum_{i\in\{x,y,z\}} \theta_i\sigma_i/2$	$\sum_{i=1}^{p} \theta_i\sigma_z^{(i)}/2$	$\sum_{i=1}^{p} \theta_i \vert i\rangle\langle i\vert \otimes \sigma_z/2$	$\sum_{i=1}^{p} \theta_i \vert i\rangle\langle i\vert$	
$N = kn$ probs.	separately	π^2	$8\pi^2 \leq ?$	$27\pi^2 \leq ?$	$\pi^2 p^2$	$\pi^2 p^3$	$\pi^2 p^3$	$\frac{1}{N^2}$
	jointly	–	$2.34\pi^2$	$4\pi^2$	$\pi^2 p$	$c_1 p^3$	$c_2 p^3$	
n probs. k reps.	separately	1	4	9	p	p^2	p^2	$\frac{1}{kn^2}$
	jointly	–	4 (2)	9 (3)	p	p^2	$p^2/4$	

multiplied by a positive-semidefinite weight matrix under the trace (which makes also nondiagonal elements of the covariant matrix important). Note, however, that such a general case can be reduced to the above after proper reparametrization. Indeed, if for some parametrization $\boldsymbol{\theta}'$ the cost is given by $\mathrm{Tr}(W\Sigma')$ (with $W \geqslant 0$), one may take A satisfying $W = A^T A$ and then $\mathrm{Tr}(W\Sigma') = \mathrm{Tr}(A^T A\Sigma') = \mathrm{Tr}(A\Sigma' A^T) = \mathrm{Tr}(\Sigma)$ (where in the last step we applied $\boldsymbol{\theta} = A\boldsymbol{\theta}'$). Therefore, for simplicity of further formulas we will only consider the cost of the form as given in Eq. (5).

We will consider multiple use of the channel. It may be seen as the action of N gates in parallel on an arbitrary entangled N-probe state:

$$\rho_{N,\boldsymbol{\theta}} = \mathcal{E}_{\boldsymbol{\theta}}^{\otimes N}(\rho_{\mathrm{in}}) \tag{6}$$

or, even more general, as a general adaptive scheme, where we apply N sequential usage of the channel, where arbitrary large ancilla and arbitrary unitary controls between the actions of the channel are allowed [1,23,68]:

$$\rho_{N,\boldsymbol{\theta}} = V_N \circ (\mathcal{E}_{\boldsymbol{\theta}} \otimes \mathbb{1}) \dots V_1 \circ (\mathcal{E}_{\boldsymbol{\theta}} \otimes \mathbb{1})(\rho_{\mathrm{in}}) \tag{7}$$

(where $V_i \circ \rho$ is a shortcut for $V_i \rho V_i^\dagger$). Note that the first one may be simulated by the latter. In such a formulation, the amount of resources corresponds to the number of uses of the channel.

The potential advantage of measuring p parameters jointly vs separately was extensively discussed in the literature with different approaches, where alternatively the amount of resources used in a single trial [45,46,62–64] or the total number of trials [65–67] was divided between parameters in separate strategy; another analysis free of the resource allocation problem was also performed [43,44] (see Appendix A for broader discussion). In this paper, we would like to focus on the situation when the same task (i.e., estimation of the set of parameters with a given cost function) is performed with a joint or a separate strategy, when the same amount of resources is used in the end. The problem is schematically shown in Fig. 1, where Bob sends to Alice N copies of a quantum gate depending on unknown parameters $\boldsymbol{\theta}$ (where each copy can only be used once) and expect from Alice that she will send him back estimated values of parameters $\tilde{\boldsymbol{\theta}}$, in a way that the total cost is minimized.

Alice may alternatively use N gates to measure all the parameters jointly or separately. In the first case, labeled as JNT, the minimal achievable cost is given by

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\mathrm{JNT}} = \min_{N \text{ protocol}} \mathrm{Tr}(\Sigma), \tag{8}$$

where by minimization over "N protocol" we understand the minimization over the choice of the initial state ρ_{in}, the unitaries V_i acting between N usage of the gate $\mathcal{E}_{\boldsymbol{\theta}}$, the measurement $\{M_x\}_x$, and the estimator $\tilde{\boldsymbol{\theta}}(x)$.

In the second case, labeled as SEP, Alice needs to divide all N gates between her "minions," sending to each of them N_i gates and ordering to focus on the measurement of a single parameter θ_i. Note that we consider the case when all the parameters have some fixed (unknown) values which do not fluctuate themselves. Hence, in the separate scenario, different parameters are estimated from independent measurements and the resulting covariance matrix will always be diagonal (assuming that all estimators are unbiased):

$$\Sigma = \mathrm{diag}(\Delta^2\tilde{\theta}_1, \dots, \Delta^2\tilde{\theta}_p). \tag{9}$$

The minimal achievable cost for such a strategy is given by

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\mathrm{SEP}} = \min_{\{N_i\}} \left(\sum_{i=1}^{p} \min_{N_i \text{ protocol}} \Delta^2\tilde{\theta}_i \right). \tag{10}$$

However, this strategy may by further optimized (while retaining the key feature that each of the minions measures only a single parameter). Indeed, Alice may demand from them to measure arbitrary linearly independent combinations of original parameters $\theta_i' = [A^{-1}\boldsymbol{\theta}]_i$ (where A is invertible matrix):

$$\Sigma' = \mathrm{diag}(\Delta^2\tilde{\theta}_1', \dots, \Delta^2\tilde{\theta}_p'), \quad \Sigma = A\Sigma' A^T. \tag{11}$$

To distinguish this strategy from the one where parametrization is fixed, we will label it as SEP+. Note that the resulting

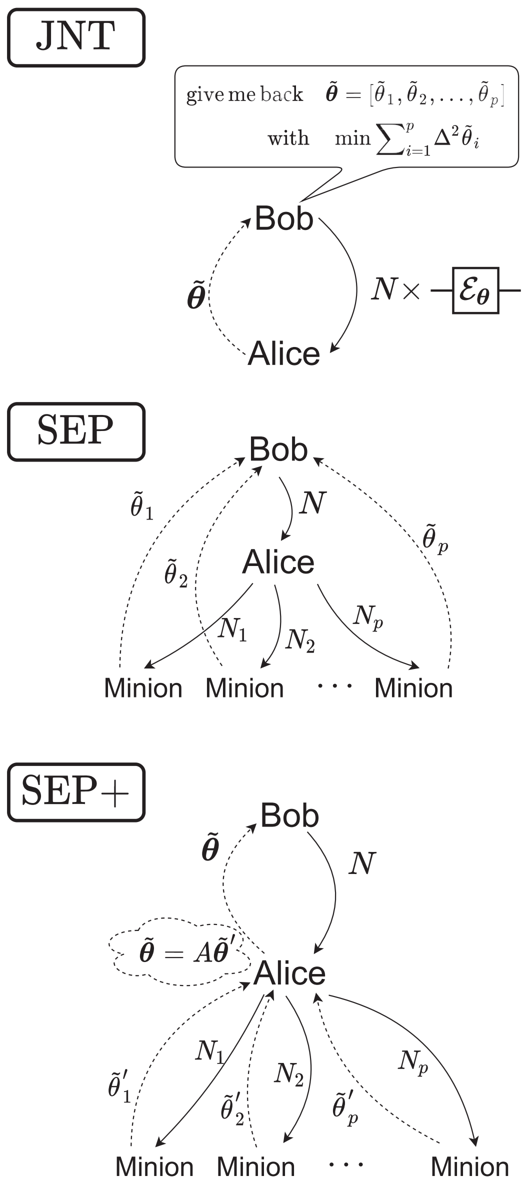

FIG. 1. Bob sends to Alice N quantum gates that depend on p unknown parameters θ_i. Her goal is to send him back their estimated values $\tilde{\theta}_i$, in order to minimize the total quadratic cost. In [JNT] she is allowed to perform full optimization and measure all the parameters jointly. In [SEP] she divides the gates between her minions, where each of them is told to measure a single parameter θ_i. In [SEP+] the minions are allowed to measure mutually linearly independent combinations $\theta'_i = [A^{-1}\boldsymbol{\theta}]_i$ and send the result back to Alice, who reconstructs the initial parameters in postprocessing. Note that [SEP+] may be seen as a special case of [JNT], while [SEP] as a special case of [SEP+].

covariance matrix is diagonal in $\boldsymbol{\theta}'$ parametrization, but not necessarily in the initial one. The resulting cost is given by

$$\sum_{i=1}^{p} \Delta^2\tilde{\theta}_i = \mathrm{Tr}(\Sigma) = \mathrm{Tr}(A\Sigma' A^T) = \sum_{i=1}^{p} [A^T A]_{ii} \Delta^2\tilde{\theta}'_i, \quad (12)$$

and therefore

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\text{SEP+}} = \min_{A,\{N_i\}} \left(\sum_{i=1}^{p} [A^T A]_{ii} \min_{N_i \text{ protocol}} \Delta^2\tilde{\theta}'_i \right). \quad (13)$$

Note that for orthogonal transformations $A = O$ where $O^T O = \mathbb{1}$, no additional term appears in Eq. (13). However, it may happen that the optimal separate strategy indeed requires a nonorthogonal transformation (see Appendix B for an example). Therefore, in general we have

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\text{JNT}} \leqslant \Delta^2\tilde{\boldsymbol{\theta}}_{\text{SEP+}} \leqslant \Delta^2\tilde{\boldsymbol{\theta}}_{\text{SEP}}, \quad (14)$$

as for each inequality the right-hand strategy may be seen as the special case of the one on the left side.

We now want to analyze these strategies further using the two paradigms: the one when the constraint is imposed on the number of gates used in a single trial n, and where the number of trials $k \gg n$, and the second one where only the total number of gates N is limited.

A. Resource distribution in separate strategies

Let us start with some general analysis of optimal resource distribution in separate strategies, focusing on the limit of large N. Assume that for each parameter θ_i, the minimal variance obtainable with the use of N_i gates (where $\sum_{i=1}^{p} N_i = N$) scales, in the leading terms, like $N_i^{-\alpha}$:

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\text{SEP}} = \min_{\{N_i\}} \sum_{i=1}^{p} \min_{N_i \text{ protocol}} \Delta^2\tilde{\theta}_i = \min_{\{N_i\}} \sum_{i=1}^{p} \frac{c_i}{N_i^{\alpha}} + o(N_i^{-\alpha}), \quad (15)$$

where all c_i have finite positive values. Neglecting term $o(N_i^{-\alpha})$, by applying the standard Lagrange multiplayers method, we obtain the optimal resources redistribution to be

$$N_i = N \frac{c_i^{1/(\alpha+1)}}{\sum_j c_j^{1/(\alpha+1)}}, \quad (16)$$

which leads to

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\text{SEP}} = \frac{1}{N^{\alpha}} \left(\sum_{i=1}^{p} c_i^{1/(\alpha+1)} \right)^{\alpha+1} + o(N^{-\alpha}). \quad (17)$$

Further in this paper, we focus only on the asymptotic behavior and for simplicity of the formulas we will omit the term $o(N^{-\alpha})$, using the sign "$\simeq$" instead. Introducing $\bar{c}_\alpha$ for the proper power mean $\bar{c}_\alpha = (\frac{1}{p}\sum_{i=1}^{p} c_i^{1/(\alpha+1)})^{\alpha+1}$ we get

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\text{SEP}} \simeq \frac{\bar{c}_\alpha p^{\alpha+1}}{N^{\alpha}}, \quad (18)$$

so (for models in which $\bar{c}_\alpha$ does not scales with p) we see a $p^{\alpha+1}$ scaling of the cost with the number of parameters involved.

Moreover, the optimal SEP and SEP+ strategies may be bounded from above by

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\text{SEP+}} \leqslant \Delta^2\tilde{\boldsymbol{\theta}}_{\text{SEP}} \lesssim p^{\alpha} \Delta^2\tilde{\boldsymbol{\theta}}_{\text{JNT}}, \quad (19)$$

as one may always consider a suboptimal separate strategy, where each of the minions performs a measurement corresponding to the optimal JNT protocol (with N/p gates), but sends to Alice only the estimated value corresponding to a single parameter θ_i.

B. Many-repetition scenario (n gates, k trials): Cramer-Rao–type bounds

Consider now the situation where the number of gates used in a single trial is restricted to some large but finite number n. Then, total amount of resources is divided into k trials, satisfying $N = nk$, and the limit $N \to \infty$ corresponds to $k \to \infty$, $n =$ const. Therefore, in the context of the discussion from the previous section, the scaling of the cost with the total amount of resources N is always linear here (as it scales linearly with k), no matter how it depends on n.

Let us briefly remind the foundations of the CR bound approach. The CR bound is based on the idea of local unbiasedness. We consider only estimators satisfying

$$\int dx\, p_{\boldsymbol{\theta}_0}(x)\tilde{\theta}_i(x) = \theta_{0i}, \quad \int dx \frac{dp_{\boldsymbol{\theta}}(x)}{d\theta_i}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tilde{\theta}_j(x) = \delta_{ij}. \tag{20}$$

For a given output state $\rho_{\boldsymbol{\theta}}$ the covariance matrix is bounded by quantum CR inequality

$$\Sigma \geqslant \frac{1}{k}F^{-1}, \quad F_{ij} = \mathrm{Tr}\left[\rho_{\boldsymbol{\theta}_0}\frac{1}{2}(L_iL_j + L_jL_i)\right], \tag{21}$$

where the matrix inequality means that $\Sigma - \frac{1}{k}F^{-1}$ is positive semidefinite and L_i are the symmetric logarithmic derivatives satisfying $\frac{d\rho_{\boldsymbol{\theta}}}{d\theta_i}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \frac{1}{2}(L_i\rho_{\boldsymbol{\theta}_0} + \rho_{\boldsymbol{\theta}_0}L_i)$. This leads to

$$\mathrm{Tr}(\Sigma) \geqslant \frac{1}{k}\mathrm{Tr}(F^{-1}). \tag{22}$$

Since measurements optimal for different parameters might be mutually incompatible, the above inequality is asymptotically saturable for large k if and only if [44]

$$\mathrm{Im}\mathrm{Tr}(\rho_{\boldsymbol{\theta}_0}L_iL_j) = 0. \tag{23}$$

More precisely, if Eq. (23) is satisfied and additionally $\rho_{\boldsymbol{\theta}}$ is pure, then there exists a local measurement (performed on a single copy of $\rho_{\boldsymbol{\theta}}$) and an estimator depending on k measurement results $\tilde{\boldsymbol{\theta}}(x_1, x_2, \dots, x_k)$, which is asymptotically unbiased and saturates Eq. (22). However, if $\rho_{\boldsymbol{\theta}}$ is the mixed state, in may in general happen that a collective measurement on all k copies of the state $\rho_{\boldsymbol{\theta}}^{\otimes k}$ is required in order to saturate the CR bound [50,52]. Therefore, the minimal achievable cost obtainable with k trials, where in each trial n gates are used, may be bounded from below by the right-hand side of Eq. (22) minimized over all feasible output states $\rho_{n,\boldsymbol{\theta}}$ [i.e., states that can be obtained when optimizing the protocol over ρ_{in} and all $V_1, \dots, V_n$ in Eq. (7)]:

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\mathrm{JNT}}^{\mathrm{CR}} \geqslant \frac{1}{k}\min_{\rho_{n,\boldsymbol{\theta}}}\mathrm{Tr}(F^{-1}), \tag{24}$$

where the inequality is tight iff Eq. (23) holds; in fact, in all examples discussed in this paper, this will be the case.

Now let us consider a separate strategy and let $\rho_{n,\boldsymbol{\theta}}^i$ be the output state designed to estimate the value of parameter θ_i. Note that in principle it may by inefficient or even impossible to fully isolate the dependence of the state on the remaining parameters $\theta_{j\neq i}$: their variations may affect the measurement results, and they cannot be omitted in the analysis of the optimal separate protocol. That means that the remaining parameters should in general be treated as nuisance parameters [69,70]. Denoting by F_i the QFI matrix corresponding to $\rho_{n,\boldsymbol{\theta}}^i$, the minimal variance of estimating parameter θ_i using k_i repetitions of an experiment is bounded by

$$\Delta^2\tilde{\theta}_i \geqslant \frac{1}{k_i}[F_i^{-1}]_{ii}. \tag{25}$$

Strictly speaking, it is not necessary for the whole matrix F_i to be invertible: it is enough if $\lim_{\epsilon\to 0^+}[(F_i + \epsilon\mathbb{1})^{-1}]_{ii}$ converges to a finite value. For example, if F_i has a block-diagonal structure, it is enough if only the block containing $[\cdot]_{ii}$ element is invertible.

Note that in general $[F_i^{-1}]_{ii} \geqslant [F_i]_{ii}^{-1}$, where the right-hand side of the inequality corresponds to a direct application of the single-parameter estimation theorem, neglecting the role of nuisance parameters. Moreover, Eq. (25) is always saturable for large k_i if collective measurements allowed [[52], Sec. 2.7], and, therefore, Eq. (17) (with $\alpha = 1$) takes the form

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\mathrm{SEP}}^{\mathrm{CR}} \simeq \frac{1}{k}\left(\sum_{i=1}^{p}\sqrt{\min_{\rho_{n,\boldsymbol{\theta}}^i}[F_i^{-1}]_{ii}}\right)^2. \tag{26}$$

After optimization over the parametrization is performed we get

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\mathrm{SEP+}}^{\mathrm{CR}} \simeq \frac{1}{k}\min_A\left(\sum_{i=1}^{p}\sqrt{[A^TA]_{ii}\min_{\rho_{n,\boldsymbol{\theta}}^i}[A^{-1}F_i^{-1}A^{-1T}]_{ii}}\right)^2, \tag{27}$$

where the rule for the transformation of the QFI matrix $F_i' = A^TF_iA$ has been used.

Using Eqs. (14) and (19) we can write

$$\Delta^2\tilde{\boldsymbol{\theta}}_{\mathrm{JNT}}^{\mathrm{CR}} \leqslant \Delta^2\tilde{\boldsymbol{\theta}}_{\mathrm{SEP+}}^{\mathrm{CR}} \leqslant \Delta^2\tilde{\boldsymbol{\theta}}_{\mathrm{SEP}}^{\mathrm{CR}} \lesssim p\Delta^2\tilde{\boldsymbol{\theta}}_{\mathrm{JNT}}^{\mathrm{CR}}. \tag{28}$$

C. Heisenberg limit (total N gates): Asymptotic local minimax bound

The methods discussed in the previous section cannot be used in a situation where only the total amount of gates N is constrained since we can not invoke the general CR saturability arguments which require a many-repetition scenario. In the single-parameter case the problem has been discussed within the MM [27] and the Bayesian [32] approaches.

In this paper we will follow the first one, as it is conceptually and technically simpler to apply in the case of estimating values of a parameter in a vicinity of a single point. Let us start by briefly reminding the idea of local asymptotic MM approach applied to single-parameter estimation [27,71].

Instead of invoking the property of local unbiasedness (as is done in the CR-based approach), we assume that the true value of the parameter lies in some finite-size neighborhood of θ_0, named $\Theta(\theta_0, \delta) = [\theta_0 - \delta/2, \theta_0 + \delta/2]$. Then, we consider a strategy, which minimizes the cost in the most pessimistic scenario (we always choose the point in Θ, where the strategy works the worst). Here, unlike in the previously discussed approach, only a single realization of the measurement is considered, and any measurement outcome is directly related with a given value of the estimator $\tilde{\theta}$. Due to this fact, and for the simplicity of notation, we may label the measurement's outcomes by $\tilde{\theta}$, so that the formula for the MM bound takes the form

$$\inf_{\rho_\theta, M_{\tilde{\theta}}} \sup_{\theta\in\Theta(\theta_0,\delta)} \int d\tilde{\theta}\, \mathrm{Tr}(M_{\tilde{\theta}}\rho_\theta)(\tilde{\theta}-\theta)^2. \quad (29)$$

This value, however, depends on the size of Θ. In order to get rid of this dependence, and hence be able to compare the results with the CR-based approach (which is effectively a single-point estimation approach), we do the following construction. Let $\{\rho_{N,\theta}, M_{N,\tilde{\theta}}\}$ be a sequence of output states (for an N-gate protocol) and the corresponding measurements. Then, assuming that for a large N the corresponding cost scales like $1/N^\alpha$ (where $\alpha = 1$ corresponds to the standard scaling, while $\alpha = 2$ to the Heisenberg scaling), define [[27], Sec. 5]

$$\mathcal{C}^{(\alpha)}[\{\rho_{N,\theta}, M_{N,\tilde{\theta}}\}] = \lim_{\delta\to 0}\lim_{N\to\infty} N^\alpha \sup_{\theta\in\Theta(\theta_0,\delta)} \times \int d\tilde{\theta}\, \mathrm{Tr}(M_{N,\tilde{\theta}}\rho_{N,\theta})(\tilde{\theta}-\theta)^2. \quad (30)$$

Note that the order of taking the limits matters (as for the opposite order the trivial constant estimator pointing θ_0 independent on the measurement results would lead to a zero cost). Intuitively, such an order approximates a situation, where one considers a δ-independent measurement strategy and for each δ checks its validity only for N which is much larger than the inverse of δ. Finally, taking the limit $\delta \to 0$ makes the results independent of the fact that the estimation around different points in Θ may be in principle harder than around θ_0 (i.e., even the value of the maximal QFI may depend on the value of θ, as for example in [46]).

It was shown in [[27], Sec. 5] that if for a given channel the Heisenberg scaling is not achievable ($\alpha = 1$), then indeed

$$\inf_{\{\rho_{N,\theta}, M_{N,\tilde{\theta}}\}} \mathcal{C}^{(1)} = \lim_{N\to\infty}\inf_{\rho_{N,\theta}} \frac{N}{F}. \quad (31)$$

Therefore, the MM (in the limit $\delta \to 0$) and the CR-based approaches return consistent results. However, if the channel estimation problem admits the Heisenberg scaling ($\alpha = 2$), then

$$\inf_{\{\rho_{N,\theta}, M_{N,\tilde{\theta}}\}} \mathcal{C}^{(2)} \geqslant \lim_{N\to\infty}\inf_{\rho_{N,\theta}} \frac{N^2}{F}, \quad (32)$$

where the inequality is not tight in general (see the next section). In this paper we are focusing only on the case $\alpha = 2$, and hence in what follows for a more compact notation we will drop the upper index.

For a general multiparameter estimation problem, let us define $\Theta(\boldsymbol{\theta}_0, \delta) = [\theta_{0,1} - \delta/2, \theta_{0,1} + \delta/2] \times \cdots \times [\theta_{0,p} - \delta/2, \theta_{0,p} + \delta/2]$ and

$$\mathcal{C}_i\big[\big\{\rho^i_{N_i,\boldsymbol{\theta}}, M_{N_i,\tilde{\theta}_i}\big\}\big] = \lim_{\delta\to 0}\lim_{N_i\to\infty} N_i^2 \sup_{\boldsymbol{\theta}\in\Theta(\boldsymbol{\theta}_0,\delta)} \times \int d\tilde{\theta}_i \mathrm{Tr}\big(M_{N_i,\tilde{\theta}_i}\rho^i_{N_i,\boldsymbol{\theta}}\big)(\tilde{\theta}_i - \theta_i)^2. \quad (33)$$

Then, for a large N, Eq. (17) (with $\alpha = 2$) up to the leading term in N takes the form

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP}} \simeq \frac{1}{N^2}\left(\sum_{i=1}^p \sqrt[3]{\inf_{\{\rho_{N_i,\boldsymbol{\theta}}, M_{N_i,\tilde{\theta}_i}\}} \mathcal{C}_i}\right)^3. \quad (34)$$

After the optimization over reparametrizations it reads as

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP+}} \simeq \frac{1}{N^2}\min_A\left(\sum_{i=1}^p \sqrt[3]{[A^TA]_{ii}\inf_{\{\rho_{N_i,\boldsymbol{\theta}'}, M_{N,\tilde{\theta}'_i}\}} \mathcal{C}'_i}\right)^3, \quad (35)$$

where $\mathcal{C}'_i$ is given by Eq. (33), after making the substitution $\theta_i \to \theta'_i = [A^{-1}\boldsymbol{\theta}]_i$. Analogously, for the joint estimation case

$$\mathcal{C}_{\mathrm{JNT}}[\{\rho_{N,\boldsymbol{\theta}}, M_{N,\tilde{\boldsymbol{\theta}}}\}] = \lim_{\delta\to 0}\lim_{N\to\infty} N^2 \sup_{\boldsymbol{\theta}\in\Theta(\boldsymbol{\theta}_0,\delta)} \times \int d\tilde{\boldsymbol{\theta}}\, \mathrm{Tr}(M_{N,\tilde{\boldsymbol{\theta}}}\rho_{N,\boldsymbol{\theta}})\mathrm{Tr}(\Sigma) \quad (36)$$

and

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \simeq \frac{1}{N^2}\inf_{\{\rho_{N,\boldsymbol{\theta}}, M_{N,\tilde{\boldsymbol{\theta}}}\}} \mathcal{C}_{\mathrm{JNT}}. \quad (37)$$

Finally, using Eqs. (14) and (19) we can write

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \leqslant \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP+}} \leqslant \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP}} \lesssim p^2\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}}. \quad (38)$$

III. HEISENBERG LIMIT BOUND

Further on we will be interested in the models where Heisenberg scaling occurs for all the estimated parameters. This is the case for noiseless unitary evolution, where the parameters enter into the evolution as multipliers of the evolution generators:

$$U_{\boldsymbol{\theta}} = e^{i\boldsymbol{\theta}\cdot\boldsymbol{\Lambda}}, \quad (39)$$

with $\boldsymbol{\Lambda} = [\Lambda_1, \ldots, \Lambda_p]^T$ where all Λ_i are mutually linearly independent.

Note that in presence of noise, Heisenberg scaling may not be achieved in general [20,72]. Still, for certain noise models, a proper quantum error correction protocol may be used to isolate the part of signal which is undisturbed by the noise and effectively obtain a purely unitary evolution [23,25,26,73] of the form (39). Hence, our discussion here will be relevant for such models as well.

A. Single-parameter saturable lower bound

As a reference point for further considerations, let us recall a paradigmatic estimation model, a single-phase estimation problem in a two-arm interferometer. In this case, the parameter encoding channel acts on the two-mode single-photon states space spanned by $\{|0\rangle, |1\rangle\}$, and is represented by

$$U_\theta = e^{i\theta|1\rangle\langle 1|}. \quad (40)$$

For simplicity, let us focus on parallel strategies, Eq. (6), first. The output state $|\psi^n_\theta\rangle = U_\theta^{\otimes n}|\psi_{\mathrm{in}}\rangle$, which maximizes the QFI, is the famous |n00n⟩ state

$$|\psi^n_\theta\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + e^{in\theta}|1\rangle^{\otimes n}), \quad (41)$$

for which $F = n^2$ and CR the bound may be saturated with protective measurement onto states $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} \pm i\,|1\rangle^{\otimes n})$. In fact, it may be also saturated by a standard photon-counting measurement and an estimator based on the parity of detected number of photons [74].

Note, however, that the state (41) is unable to distinguish between the phases that differ by a multiple of $2\pi/n$ [[27], Sec. 5]. Therefore, it only allows for estimation of the parameter in a small region of $[\theta_0 - \pi/n, \theta_0 + \pi/n]$. While many-repetition scenario is under consideration, this issue does not generate a serious problem, as even starting with the unknown phase, for $k \gg n$ one may always spend the first $\sqrt{k}$ trials to find such small region [for example, by using the product states $\frac{1}{\sqrt{2^n}}(|0\rangle + |1\rangle)^{\otimes n}$ in each repetition] and next in remaining $k - \sqrt{k}$ trials use |n00n⟩ states to finally achieve precision $1/kn^2$ (up to the leading term in k). However, it is clear that an analogous single |N00N⟩ state cannot be used to obtain the fundamental Heisenberg limit, if inserted into Eq. (30) it would lead to $\mathcal{C} = +\infty$.

Calculation of the minimal obtainable value of $\mathcal{C}$ by a direct minimization of Eq. (30) is a hard problem. Fortunately, the task may be significantly simplified by taking into account a symmetry of the problem.

First note that for such a channel the two phases which differ by a factor 2π should be regarded as equivalent. Therefore, we consider a periodic cost function of the form

$$\mathrm{cost}(\theta, \tilde{\theta}) = 4\,\sin^2\left(\frac{\theta - \tilde{\theta}}{2}\right), \tag{42}$$

which reflects this property whereas for small difference may be well approximated by $\approx (\theta - \tilde{\theta})^2$.

Next, as argued in [27], the local asymptotic MM cost [Eq. (30)] for this problem is exactly the same as the minimal obtainable cost for a completely unknown phase $\theta \in \Theta = [0, 2\pi)$ multiplied by N^2 (in the limit $N \to \infty$). Intuitively, the reason for this is that for any finite δ, when we start from a completely unknown phase, we always spend at the beginning $\sqrt{N}$ gates to discriminate the region of size δ, where the true value of θ lies (with the probability of the error decreasing as exponentially fast with N), and use the remaining $N - \sqrt{N}$ to estimate the value inside this region.

Note that while at a first glance the above construction seems to require adaptiveness (as in the second step we use the information from the first one), in may in fact be performed also within the parallel scheme (see details in Appendix C). See also [32,61,75] for further discussion about optimization of region discrimination in the first part of the above construction, as well as a general discussion about this bound for finite N.

The state which is optimal for measuring a completely unknown phase is the |SIN⟩ state [33–35]

$$\left|\psi^N_{\boldsymbol{\theta}}\right\rangle = \sum_{m=0}^{N} e^{im\theta} \frac{\sqrt{2}}{\sqrt{N+2}} \sin\left(\frac{(m+1)\pi}{N+2}\right) |N-m\rangle_0\, |m\rangle_1\,, \tag{43}$$

where $|N-m\rangle_0\, |m\rangle_1$ is a fully symmetric state with m photons in the sensing arm $|1\rangle$ and $N-m$ is the reference arm $|0\rangle$. The corresponding mean cost obtainable with applying covariant measurement is

$$\forall_\theta \quad \int d\tilde{\theta}\; \mathrm{Tr}(M_{N,\tilde{\theta}}\rho_{N,\theta})\mathrm{cost}(\theta, \tilde{\theta}) = 2\left[1 - \cos\left(\frac{\pi}{N+2}\right)\right] \tag{44}$$

and therefore the constant which multiplies the leading term $1/N^2$ equals

$$\inf_{\{\rho_{N,\boldsymbol{\theta}}, M_{N,\tilde{\theta}}\}} \mathcal{C} = \lim_{N\to\infty} N^2 2\left[1 - \cos\left(\frac{\pi}{N+2}\right)\right] = \pi^2. \tag{45}$$

The following natural question arises: Is there a simple interpretation of this π^2 factor discrepancy between the minimal achievable variance and the inverse of maximal Fisher information? In fact, one may indeed use the |n00n⟩ states even when estimating a completely unknown phase, provided one divides all the available resources N into M subsets, each of size m_i and then use M states of the form $\frac{1}{\sqrt{2}}(|0\rangle^{\otimes m_i} + |1\rangle^{\otimes m_i})$ for estimation. It was shown numerically that for the optimal distribution of m_i, the overhead factor π^2 is indeed recovered [76,77]. Some suboptimal strategies have also been demonstrated experimentally, revealing only a slightly bigger variance [78,79].

The above result may be directly generalized for an arbitrary quantum channel of the form $U_\theta = e^{i\theta\Lambda}$ [32]. Denoting by $\lambda[\Lambda]$ the difference between the maximal and the minimal eigenvalues of the operator Λ, for optimal usage of N quantum gates we have

$$\inf_{\{\rho_{N,\boldsymbol{\theta}}, M_{N,\tilde{\theta}}\}} \mathcal{C} = \frac{\pi^2}{\lambda^2[\Lambda]} \Rightarrow \Delta^2\tilde{\theta}^{\mathrm{MM}} \simeq \frac{\pi^2}{N^2\lambda^2[\Lambda]}. \tag{46}$$

For comparison, in the scenario, where $k \to \infty$ repetitions is considered (with usage n gates in each of them), minimal obtainable cost is given by

$$\max_{\rho_{n,\theta}} F = n^2\lambda^2 \Rightarrow \Delta^2\tilde{\theta}^{\mathrm{CR}} \simeq \frac{1}{kn^2\lambda^2[\Lambda]}. \tag{47}$$

In both scenarios, this optimal precision is obtainable already in the parallel scheme and it cannot be beaten by any adaptive protocol that involves additional action of V_i operations in-between [32].

B. Multiparameter unitary estimation lower bound

Consider a general problem of local unitary channel estimation

$$U_{\boldsymbol{\theta}} = e^{i\boldsymbol{\theta}\cdot\boldsymbol{\Lambda}}, \tag{48}$$

with linearly independent generators Λ_i acting on a d-dimensional space. First note that for such a formulated problem the Heisenberg scaling is indeed achievable. This was shown for the most general SU(d) estimation problem for any d using a parallel scheme, in both many-repetition scenario [80] (where the exact fundamental bound has been derived and proven to be saturable if entanglement with ancilla allowed) and single-repetition scenario [60] (where optimal scaling has been proven to be $\propto 1/N^2$ with an exemplary state satisfying this scaling). As for any finite-dimensional space, the problem states in Eq. (48) may be seen as estimation of some subset of SU(d) generators, the statement is proven.

Let us first argue that from the point of view of estimation of any single parameter θ_i, the existence of an additional part of the generator $\sum_{j\neq i}\theta_j\Lambda_j$ cannot help in estimation, i.e., it cannot decrease the minimal achievable cost. Note that a single gate $U_{\boldsymbol{\theta}}$ (where all generators act jointly) may be arbitrarily well approximated (for a sufficiently large l) by

$$U_{\boldsymbol{\theta}} = e^{i\boldsymbol{\theta}\cdot\boldsymbol{\Lambda}} \overset{l\gg 1}{\approx} \left(e^{i\theta_i\frac{\Lambda_i}{l}} e^{i\sum_{j\neq i}\theta_j\frac{\Lambda_j}{l}}\right)^l \tag{49}$$

(more precisely, the above approximation is exact in the limit $l\to\infty$ due to the Trotter formula [81]). Therefore, an N-fold action of $U_{\boldsymbol{\theta}}$ may be seen as lN 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. After such a procedure the product $\lambda[\Lambda_i/l](lN) = \lambda[\Lambda_i]N$ remains unchanged. Therefore, both the asymptotic value of the bound and the rate of its convergence remain the same. Consequently, for each θ_i we have

$$\inf_{\{\rho^i_{N_i,\boldsymbol{\theta}},M_{N_i,\tilde{\theta}}\}} \mathcal{C}_i \geqslant \frac{\pi^2}{\lambda^2[\Lambda_i]}. \tag{50}$$

Note, however, that the saturability of the bound is not guaranteed. Let us now bound from below the minimal achievable cost by assuming the most optimistic scenario, the one where not only the existence of the other part of the Hamiltonian $\sum_{j\neq i}\theta_j\Lambda_j$ does not disturb the sensing of θ_i, but also that there exists a single input state and a single measurement which are simultaneously optimal for sensing of all the parameters. Then, for any finite N we have

$$\min_{\rho_{N,\boldsymbol{\theta}},M_{N,\tilde{\boldsymbol{\theta}}}} \mathrm{Tr}(\Sigma) = \min_{\rho_{N,\boldsymbol{\theta}},M_{N,\tilde{\boldsymbol{\theta}}}} \sum_{i=1}^p \Delta^2\tilde{\theta}_i \geqslant \sum_{i=1}^p \min_{\rho^i_{N,\boldsymbol{\theta}},M_{N,\tilde{\theta}_i}} \Delta^2\tilde{\theta}_i \tag{51}$$

and, therefore, taking the asymptotic limit $N\to\infty$, we can write

$$\inf_{\{\rho_{N,\boldsymbol{\theta}},M_{N,\tilde{\boldsymbol{\theta}}}\}} \mathcal{C}_{\mathrm{JNT}} \geqslant \sum_{i=1}^p \inf_{\{\rho^i_{N,\boldsymbol{\theta}},M_{N,\theta_i}\}} \mathcal{C}_i \geqslant \sum_{i=1}^p \frac{\pi^2}{\lambda^2[\Lambda_i]}. \tag{52}$$

Notice that we have some freedom in choosing the parametrization in Eq. (48). Indeed, all the steps remain valid after an application of any orthogonal rotation in the parameter space $\boldsymbol{\theta}' = O^{-1}\boldsymbol{\theta}$, $\boldsymbol{\Lambda}' = O^T\boldsymbol{\Lambda}$, which does not change the local cost function. Note, however, that the restriction to orthogonal transformations $O^TO = \mathbb{1}$ is crucial here, as for a more general one A the off-diagonal elements may appear in the formula for the cost $\mathrm{Tr}(A^TA\Sigma)$, which would make it impossible to bound it using Eq. (51). Therefore, we can tighten the bound resulting from Eq. (52) and write

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \gtrsim \max_O \frac{1}{N^2}\sum_{i=1}^p \frac{\pi^2}{\lambda^2([O^T\boldsymbol{\Lambda}]_i)}. \tag{53}$$

An analogous bound may be derived for the trace of the inverse of the QFI (see also [26]):

$$\mathrm{Tr}(F^{-1}) = \sum_{i=1}^p [F^{-1}]_{ii} \geqslant \sum_{i=1}^p [F_i]^{-1}_{ii} \geqslant \sum_{i=1}^p \left[F_i^{-1}\right]_{ii} = \sum_{i=1}^p \frac{1}{n^2\lambda_i^2} \tag{54}$$

so

$$\mathrm{Tr}(F^{-1}) \geqslant \max_O \frac{1}{n^2}\sum_{i=1}^p \frac{1}{\lambda^2([O^T\boldsymbol{\Lambda}]_i)} \tag{55}$$

and

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT}} \gtrsim \max_O \frac{1}{kn^2}\sum_{i=1}^p \frac{1}{\lambda^2([O^T\boldsymbol{\Lambda}]_i)}, \tag{56}$$

where the same O maximizes both Eqs. (53) and (56).

C. Relation between optimal global and local minimax costs

It is worth mentioning that in the cases where the local estimation problem may be extended to the covariant group estimation problem, the local minimax cost is the same as the one for estimating a completely unknown element of the group. Let us formalize it.

Let U_g (where $g\in G$) be a unitary representation of compact group G (so $U_{g_1}U_{g_2} = U_{g_1g_2}$) and consider the cost function invariant with respect to the action of this group $\forall_{g,\tilde{g},h\in G}\mathrm{cost}(hg,h\tilde{g}) = \mathrm{cost}(g,\tilde{g})$. Let $\boldsymbol{\theta} = [\theta_1,\ldots,\theta_p]\mapsto g_{\boldsymbol{\theta}}$ be a local parametrization around neutral element of the group $e\in G$ such that $U_{g_{\boldsymbol{\theta}}} = e^{i\boldsymbol{\theta}\boldsymbol{\Lambda}}$. Then assuming that $\mathrm{cost}(e,g_{\tilde{\boldsymbol{\theta}}}) = \tilde{\boldsymbol{\theta}}^2 + o(\tilde{\theta}_i\tilde{\theta}_j)$, the asymptotic minimax cost for $\mathrm{cost}(g,\tilde{g})$ for $g\in G$ is the same as the local one, Eq. (36) (see Appendix C for more details). The reasoning is based on the same idea as the one performed for single-parameter case [27] (reminded here in Sec. III A).

Moreover, in [82] it was shown that for covariant estimation the optimal results may be obtained within a parallel scheme (without the necessity of involving adaptiveness), which implies that also in the local minimax approach there is no advantage in applying adaptive strategy (in contrast to the results obtainable within CR formalism, which will be discussed in Sec. IV C).

D. Separate strategy lower bound

Finally, we would like to derive a simple bound for the minimal cost obtainable by the SEP+ strategy, which will allow us for a quick assessment of potential benefits due to a rotation in the parameter space. We have

$$\begin{aligned}\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP+}} &\gtrsim \min_{A,N_j}\sum_{j=1}^p \frac{\pi^2}{N_j^2}\frac{[A^TA]_{jj}}{\lambda^2([A^T\boldsymbol{\Lambda}]_j)}\\ &\geqslant \min_{A,N_j}\sum_{j=1}^p \frac{\pi^2}{N_j^2}\left(\min_i \frac{[A^TA]_{ii}}{\lambda^2([A^T\boldsymbol{\Lambda}]_i)}\right)\\ &= \frac{p^3\pi^2}{N^2}\min_{A,i}\frac{[A^TA]_{ii}}{\lambda^2([A^T\boldsymbol{\Lambda}]_i)}.\end{aligned} \tag{57}$$

Note that the last term in above inequality depends only on the ith column of A. Therefore, minimization over both A and i is equivalent to minimization over a single vector $\boldsymbol{a}$:

$$\begin{aligned}\min_{A,i}\frac{[A^TA]_{ii}}{\lambda^2([A^T\boldsymbol{\Lambda}]_i)} &= \min_{\boldsymbol{a}}\frac{|\boldsymbol{a}|^2}{\lambda^2[\boldsymbol{a}\boldsymbol{\Lambda}]}\\ &= \min_{\boldsymbol{a}:|\boldsymbol{a}|^2=1}\frac{1}{\lambda^2[\boldsymbol{a}\boldsymbol{\Lambda}]} = \frac{1}{\max_{\boldsymbol{a}:|\boldsymbol{a}|^2=1}\lambda^2[\boldsymbol{a}\boldsymbol{\Lambda}]},\end{aligned} \tag{58}$$

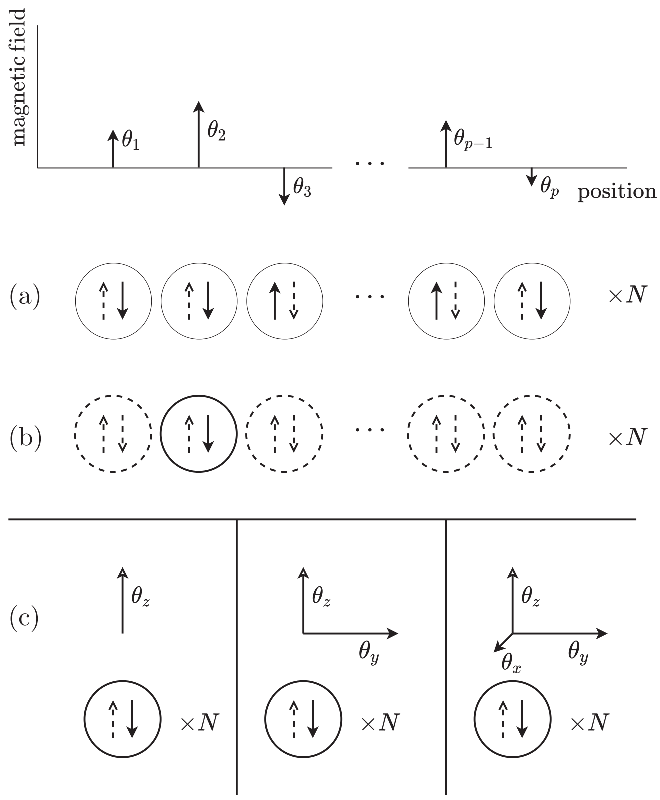

FIG. 2. Various models of magnetic field sensing by spin-$\frac{1}{2}$ atoms are discussed. In (a) and (b) the spatial distribution of magnetic field oriented in the z direction is to be estimated. In (a) the atoms are uniformly distributed in p points and one has the freedom to choose their spin orientations in the optimal way, while in (b) for every single atom both the position and the spin orientation may be chosen arbitrary. In (c) the magnetic field in single point is measured, but multicomponent estimation is discussed.

hence,

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{MM}}_{\text{SEP+}} \gtrsim \frac{p^3\pi^2}{N^2}\frac{1}{\max_{\boldsymbol{a}:|\boldsymbol{a}|^2=1}\lambda^2[\boldsymbol{a}\boldsymbol{\Lambda}]}. \tag{59}$$

Similarly, in the multiple-repetition scenario we could write (see also [26])

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{CR}}_{\text{SEP+}} \gtrsim \frac{p^2}{kn^2}\frac{1}{\max_{\boldsymbol{a}:|\boldsymbol{a}|^2=1}\lambda^2[\boldsymbol{a}\boldsymbol{\Lambda}]}. \tag{60}$$

Thanks to the bounds derived in this section, we will be able to get an insight into the benefits of joint vs separate strategies, even if we will not always be able to obtain a rigorous solution for the optimal achievable cost (see the next section).

IV. EXAMPLES

We will focus here on models which are inspired by various magnetic field sensing problems, but which are representative for a wide range of multiparameter unitary estimation problems (see Fig. 2). As shown in [61], if all the generators mutually commute $\forall_i[\Lambda_i,\Lambda_j]=0$, there is no asymptotic advantage (for large N) in using a general adaptive strategy when compared to the parallel one. Therefore, when analyzing the first two examples, we will focus on the parallel scheme only. For the last one, where the generators do not commute, both strategies will be discussed.

A. Spatially distributed magnetic field sensing: Fixed atom positions

Consider the problem of spatially distributed magnetic field sensing (which is directed along the z axis). The field is sensed by spin-$\frac{1}{2}$ atoms allocated in p spatially separated places.

Before moving on to the general solution of such a problem, we would like first to discuss it with an additional constraint imposed. Namely, we assume that the spatial distribution of sensing atoms is uniform and fixed, i.e., in each of p points there are exactly N atoms and one has only the freedom in choosing their spin orientations (this form of the problem was discussed in the many-repetition scenario in [26]). Here, by an elementary amount of resources we understand a single layer of p atoms, and the corresponding Hilbert space is the 2^p-dimensional one spanned by the vectors of the form

$$|\boldsymbol{s}\rangle = |s_1, s_2, \dots, s_p\rangle, \quad \text{with} \quad s_i\in\{-1,+1\}. \tag{61}$$

Then the elementary quantum gate is

$$U_{\boldsymbol{\theta}} = e^{i\boldsymbol{\theta}\cdot\boldsymbol{\Lambda}}, \quad \Lambda_i = \mathbb{1}^{\otimes i-1}\otimes\tfrac{1}{2}\sigma_z\otimes\mathbb{1}^{\otimes p-i}, \tag{62}$$

so $\Lambda_i|\boldsymbol{s}\rangle = \frac{1}{2}s_i|\boldsymbol{s}\rangle$. Note that here the number of quantum gates N is equal to number of p-atom layers and in fact correspond to usage of Np atoms.

Let us consider first the SEP strategy (without optimization over reparametrizations). Since $\forall_i\lambda[\Lambda_i]=1$, one gets, respectively, for the CR approach [using the analogs of $|\text{n00n}\rangle$, Eq. (41), states]

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{CR}}_{\text{SEP}} \simeq p\times\frac{1}{k/p}\times\frac{1}{n^2} = \frac{p^2}{kn^2}, \tag{63}$$

and for the MM approach [with the use of $|\text{SIN}\rangle$, Eq. (43), states]

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{MM}}_{\text{SEP}} \simeq p\times\frac{\pi^2}{(N/p)^2} = \frac{\pi^2p^3}{N^2}. \tag{64}$$

However, looking at Eqs. (59) and (60) one may see that there is a significant potential for improvement in SEP+, as the value of $\lambda[\boldsymbol{a}\boldsymbol{\Lambda}]$ will be maximal for $\boldsymbol{a}=1/\sqrt{p}[1,1,\dots,1]$ and equal to $\sqrt{p}$. Below we show a concrete reparametrization for which the mentioned bounds may be saturated. For simplicity, let us restrict to the case $p=2^r$ for some natural r.

In order to optimize the separate strategy one needs to find a parametrization for which each parameter may be sensed by all the atoms simultaneously. Therefore, instead of measuring the magnetic field point by point in p positions, one may decompose the field into a proper components by the Walsh-Hadamard transformation $\boldsymbol{\theta}'=O^{-1}\boldsymbol{\theta}$ with

$$O_{ij} = \frac{1}{\sqrt{p}}\prod_{k=0}^{r-1}(-1)^{i_kj_k}, \quad \text{where} \quad i=\sum_{k=0}^{r-1}i_k2^k \tag{65}$$

or, equivalently,

$$O = \frac{1}{\sqrt{p}}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}^{\otimes r}. \tag{66}$$

After the application of such a transformation, all the generators $[O^T\boldsymbol{\Lambda}]_i$ remain diagonal in the basis $\{|\boldsymbol{s}\rangle\}$ and moreover

$$\langle \boldsymbol{s}| [O^T\boldsymbol{\Lambda}]_i |\boldsymbol{s}\rangle = \frac{1}{2\sqrt{p}}\sum_{j=1}^{p} O_{ij}s_j. \tag{67}$$

From that, indeed, $\forall_i\ \lambda([O^T\boldsymbol{\Lambda}]_i) = \sqrt{p}$ and, moreover, as the eigenvectors of Λ_i with minimal and maximal eigenvalues take the form

$$\begin{aligned}|\lambda_{i+}\rangle &= |+O_{i1}, +O_{i2}, \dots, +O_{ip}\rangle\,,\\ |\lambda_{i-}\rangle &= |-O_{i1}, -O_{i2}, \dots, -O_{ip}\rangle\end{aligned} \tag{68}$$

the remaining generators acts on them trivially:

$$\Lambda_j\,|\lambda_{i\pm}\rangle = \pm\delta_{ij}\tfrac{1}{2}\,|\lambda_{i\pm}\rangle\,. \tag{69}$$

Hence, when focusing on the estimation of a given parameter θ_i, there are no disturbance issues related with the presence of the other parameters. Therefore,

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{CR}}_{\text{SEP+}} \simeq \frac{p}{kn^2},\quad \Delta^2\tilde{\boldsymbol{\theta}}^{\text{MM}}_{\text{SEP+}} \simeq \frac{p^2\pi^2}{N^2}. \tag{70}$$

We see that, thanks to the application of a proper reparametrization, we have decreased the cost obtainable in a separate strategy by a factor of p. Note that in order to use Eq. (65) we have assumed $p = 2^r$. However, if this is not satisfied, one can still obtain qualitatively similar results, e.g., applying parameter transformations from [83].

Going back to the initial parametrization, let us now discuss a joint strategy. As each of the parameters is associated with a different atom, all of them may be measured without disturbing the measurement outcomes of the remaining ones. More formally, as the Hilbert space corresponding to the single layer of atoms has the characteristic structure $\mathcal{H} = (\mathbb{C}^2)^{\otimes p}$, then for N layers it may be written in the form $\mathcal{H}^{\otimes N} = ((\mathbb{C}^2)^{\otimes p})^{\otimes N} = ((\mathbb{C}^2)^{\otimes N})^{\otimes p}$. Hence, we may use the $|\text{n00n}\rangle^{\otimes p}$ state (in the CR approach) or $|\text{SIN}\rangle^{\otimes p}$ state (in the MM approach), which yields

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{CR}}_{\text{JNT}} \simeq \frac{p}{kn^2},\quad \Delta^2\tilde{\boldsymbol{\theta}}^{\text{MM}}_{\text{JNT}} \simeq \frac{p\pi^2}{N^2}, \tag{71}$$

and which saturates the bound (53) (which is the tightest for the original parametrization, i.e., with $O = \mathbb{1}$).

Comparing Eq. (70) with Eq. (71), we see that for this model the existence of the advantage of measuring parameters jointly depends on the chosen paradigm: in the many-repetition scenario there is no advantage, while for the fully optimal usage of all resources the advantage increases linearly with the number of parameters.

B. Spatially distributed magnetic field sensing: Arbitrary spatial distribution of atoms

Let us now consider the same problem with the full freedom in the distribution of the atoms in both space (p positions) and spin orientations. The single-atom Hilbert space will, therefore, be spanned by

$$|i, s\rangle\,,\quad \text{with}\quad i\in\{1,2,\dots,p\},\ s\in\{-1,+1\}. \tag{72}$$

The corresponding single quantum gate has the form

$$\begin{aligned}&U_{\boldsymbol{\theta}} = e^{i\boldsymbol{\theta}\cdot\boldsymbol{\Lambda}},\\ &\Lambda_i = |i\rangle\langle i|\otimes\tfrac{1}{2}\sigma_z = \tfrac{1}{2}(|i,+\rangle\langle i,+| - |i,-\rangle\langle i,-|),\end{aligned} \tag{73}$$

so $\Lambda_i\,|j,s\rangle = \delta_{ij}\frac{s}{2}\,|j,s\rangle$. Note that here, unlike in the previous example, the number of quantum gates is equal exactly to the number of atoms used.

Similarly as in Eq. (43), without loss we may restrict to the fully symmetric space, so that any N-atomic output state may be written as

$$\left|\psi^N_{\boldsymbol{\theta}}\right\rangle = \sum_{|\boldsymbol{m}|=N} e^{\frac{i}{2}\sum\theta_i(m_{i+}-m_{i-})}c_{\boldsymbol{m}}\,|\boldsymbol{m}\rangle\,, \tag{74}$$

where $\boldsymbol{m} = [m_{1+}, m_{1-}, m_{2+}, \dots, m_{p-}]$ and $|\boldsymbol{m}| = \sum_{i=1}^{p}(m_{i+}+m_{i-})$.

In this case $\forall_i\ \lambda[\Lambda] = 1$; their nonzero eigenspaces are mutually orthogonal and hence cannot be increased by taking any linear combination $\boldsymbol{a\Lambda}$. Therefore, there is no advantage of the SEP+ over the SEP protocol, and we simply get

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{CR}}_{\text{SEP}} \simeq \frac{p^2}{kn^2} \tag{75}$$

for the state

$$\left|\psi^n_{\boldsymbol{\theta}}\right\rangle = \frac{1}{\sqrt{2}}(e^{in\theta_i/2}\,|n\rangle_{i+}\,|0\rangle_{i-} + e^{-in\theta_i/2}\,|0\rangle_{i+}\,|n\rangle_{i-}), \tag{76}$$

while in case of full optimization paradigm we get

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{MM}}_{\text{SEP}} \simeq \frac{p^3\pi^2}{N^2} \tag{77}$$

for the state

$$\begin{aligned}\left|\psi^{N/p}_{\boldsymbol{\theta}}\right\rangle = &\sum_{m=0}^{N/p}\frac{\sqrt{2}e^{i(2m-N/p)/2\theta_i}}{\sqrt{N/p+2}}\\ &\times\ \sin\left(\frac{(m+1)\pi}{N/p+2}\right)|m\rangle_{i+}\,|N/p-m\rangle_{i-}\,,\end{aligned} \tag{78}$$

where $|m\rangle_{i+}\,|N/p-m\rangle_{i-}$ denotes a state with N/p atoms in the ith position, with m of them being oriented up, and $N/p - m$ down.

Much more interesting aspects may be observed when analyzing the joint strategy. First, let us look what insight may be obtained from the bounds (56) and (53). To do so, we will use again the Walsh-Hadamard transformation. However, note that in this case the effect is completely opposite to the one observed in the previous example (application of this transformation decreases all $\lambda[O^T\Lambda_i] = 1/\sqrt{p}$), that is, in the case where one has the full freedom of distributing atoms in space, measuring combinations of magnetic fields at various points, instead of measuring the field point by point, is an inefficient separate strategy. Still, thanks to this observation, such a transformation may be used to tighten the bounds (56) and (53). Since $\sum_{i=1}^{p} 1/\lambda^2([O^T\boldsymbol{\Lambda}]_i) = p^2$, it gives

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{CR}}_{\text{JNT}} \gtrsim \frac{p^2}{kn^2},\quad \Delta^2\tilde{\boldsymbol{\theta}}^{\text{MM}}_{\text{JNT}} \gtrsim \frac{p^2}{N^2}. \tag{79}$$

In the CR case, the above bound may be saturated using the state

$$\left|\psi_{\theta}^{n}\right\rangle = \frac{1}{\sqrt{p}}\sum_{i=1}^{p}\frac{1}{\sqrt{2}}(e^{+in\theta_i/2}\left|n\right\rangle_{i,+} + e^{-in\theta_i/2}\left|n\right\rangle_{i,-}), \quad (80)$$

which is simply an equally weighted superposition of Eq. (76) and for which

$$F_{ij} = 4\,\mathrm{Re}[\langle\psi_i|\psi_j\rangle - \langle\psi_i|\psi\rangle\langle\psi|\psi_j\rangle] \Rightarrow F_{ij} = \frac{\delta_{ij}}{p} \quad (81)$$

and, consequently,

$$\mathrm{Tr}(F^{-1}) = \frac{p^2}{n^2} \Rightarrow \Delta^2\tilde{\theta}_{\mathrm{JNT}}^{\mathrm{CR}} \simeq \frac{p^2}{kn^2}. \quad (82)$$

To analyze the joint strategy within the MM approach, we will use the Fourier analysis, originally applied to the single-parameter estimation problem in [75], later generalized for another group estimation problem [59], and very recently applied for multiphase estimation in multiarm interferometer with constraining for the total number of photons [61]. Note that from the point of view of the discussed problem, only the differences between the number of atoms oriented up and down $\Delta m_i = (m_{i+} - m_{i-})/2$ matter. Therefore, for a given state (74) we define normalized states $|\Delta\boldsymbol{m}\rangle$ and coefficients $c_{\Delta\boldsymbol{m}}$ satisfying

$$\forall_{\Delta\boldsymbol{m}} c_{\Delta\boldsymbol{m}}\left|\Delta\boldsymbol{m}\right\rangle = \sum_{\boldsymbol{m}:\forall(m_{i+}-m_{i-})/2=\Delta m_i} c_{\boldsymbol{m}}\left|\boldsymbol{m}\right\rangle. \quad (83)$$

Then, Eq. (74) may be rewritten in the form

$$\left|\psi_{\theta}^{N}\right\rangle = \sum_{\sum_i|\Delta m_i|\leqslant N/2} e^{i\boldsymbol{\theta}\Delta\boldsymbol{m}} c_{\Delta\boldsymbol{m}}\left|\Delta\boldsymbol{m}\right\rangle. \quad (84)$$

Next, for N large enough, we may replace discrete variables by continuous ones $\frac{\Delta m_i}{N} \to \mu_i \in [-1/2, +1/2]$ to get

$$\left|\psi_{f,\theta}^{N}\right\rangle = \int_{\sum_i|\mu_i|\leqslant 1/2} d\boldsymbol{\mu}\, e^{iN\boldsymbol{\theta}\boldsymbol{\mu}} f(\boldsymbol{\mu})\left|\boldsymbol{\mu}\right\rangle. \quad (85)$$

As argued in [61], the optimal measurement in the asymptotic limit will be the covariant one:

$$|\chi_{\tilde{\theta}}\rangle = \frac{1}{\sqrt{(2\pi/N)^p}}\int d\boldsymbol{\mu}\, e^{iN\boldsymbol{\mu}\tilde{\boldsymbol{\theta}}}\left|\boldsymbol{\mu}\right\rangle. \quad (86)$$

For technical reasons, in further calculations we will treat the function $f(\boldsymbol{\mu})$ appearing in Eq. (85) as the one defined on the whole $\mathbb{R}^p$, but equal zero everywhere outside of $\{[\mu_1,\dots,\mu_p]\}_{\sum_i|\mu_i|\leqslant 1/2}$ (which allows us to perform the standard Fourier transform of this function). The mean value of the quadratic cost is given by

$$\int_{\mathbb{R}^p} d\tilde{\boldsymbol{\theta}}\left|\left\langle\chi_{\tilde{\theta}}\middle|\psi_{f,\theta}^{N}\right\rangle\right|^2(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})^2 = \frac{1}{N^2}\int_{\mathbb{R}^p} d\tilde{\boldsymbol{\theta}}\,|\hat{f}(\tilde{\boldsymbol{\theta}})|^2\tilde{\boldsymbol{\theta}}^2, \quad (87)$$

where $\hat{f}$ is the Fourier transform of f and we dropped the irrelevant dependence on $\boldsymbol{\theta}$.

Going back to the $\boldsymbol{\mu}$ representation and performing the minimization over f we get

$$\Delta^2\tilde{\theta}_{\mathrm{JNT}}^{\mathrm{MM}} \simeq \frac{1}{N^2}\min_f \int_{\sum_i|\mu_i|\leqslant 1/2} d\boldsymbol{\mu}\, f^*(\boldsymbol{\mu})\left(\sum_{k=1}^{p} -\partial_{\mu_k}^2\right)f(\boldsymbol{\mu}),$$
$$\text{with} \quad \int_{\sum_i|\mu_i|\leqslant 1/2} d\boldsymbol{\mu}\,|f(\boldsymbol{\mu})|^2 = 1,$$
$$f(\boldsymbol{\mu}) = 0 \quad \text{for } \boldsymbol{\mu} \text{ on the boundary } \sum_i|\mu_i| = 1/2. \quad (88)$$

The problem is therefore equivalent to minimization of the kinetic energy of a particle in infinite potential well in a shape of a p-dimensional simplex. The analytical solutions are known only for $p = 1, 2$ (see Appendix D). For higher number of parameters we will derive a lower bound. As $f(\boldsymbol{\mu}) = 0$ everywhere outside of $\sum_i|\mu_i| \leqslant \frac{1}{2}$, the mean value of $\sum_i|\mu_i|$ is trivially smaller or equal $\frac{1}{2}$. Next, thanks to the symmetry, we may assume, without loss of generality, that the function $f(\mu)$ minimizing the above is fully symmetric under the exchange of variables μ_i and therefore all the mean values of $|\mu_i|$ are equal and $\leqslant 1/(2p)$ (see also Supplemental Material of [61] for a broader discussion of this argument). The minimal sum of variances may be therefore be bounded as

$$\Delta^2\tilde{\theta}_{\mathrm{JNT}}^{\mathrm{MM}} \gtrsim p \times \min_g \int_{-\infty}^{+\infty} d\mu\, g^*(\mu)\left(-\frac{\partial^2}{\partial^2\mu}\right)g(\mu) \quad (89)$$

with

$$\int_{-\infty}^{+\infty} d\mu|g(\mu)|^2 = 1, \quad \int_{-\infty}^{+\infty} d\mu|g(\mu)|^2|\mu| = \frac{1}{2p}. \quad (90)$$

Moreover, from the symmetry of the problem, the solution will be symmetric with respect to the 0 point (which, assuming differentiability, implies $\frac{\partial g}{\partial\mu}|_{\mu=0} = 0$). Therefore, the problem is equivalent to

$$\min_g \int_0^{+\infty} d\mu\, g^*(\mu)\left(-\frac{\partial^2}{\partial^2\mu}\right)g(\mu), \quad \text{with} \quad \left.\frac{\partial g}{\partial\mu}\right|_{\mu=0} = 0,$$
$$\int_0^{+\infty} d\mu|g(\mu)|^2 = 1, \quad \int_0^{+\infty} d\mu|g(\mu)|^2|\mu| = \frac{1}{2p}, \quad (91)$$

which may be solved using the Lagrange multiplayers method. The solution in terms of the Airy function yields the final bound (see Appendix D for detail derivation)

$$\Delta^2\tilde{\theta}_{\mathrm{JNT}}^{\mathrm{MM}} \gtrsim \frac{0.63p^3}{N^2}. \quad (92)$$

We are unable to prove the tightness of the above bound. However, we are able to point out an exemplary state for which

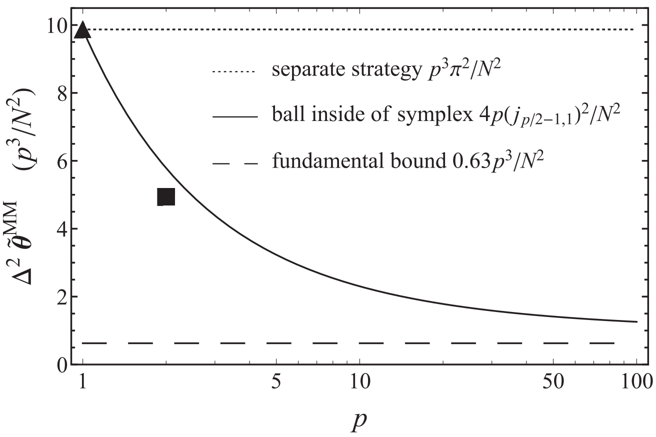

FIG. 3. The minimal achievable cost in the problem of estimating spatial distribution of magnetic field in p point is analyzed in the limit of large N (with no repetition scenario). The triangle and square represent analytically found results for $p = 1, 2$. The solid line represents the variance obtainable by exemplary suboptimal state state. The dotted line corresponds to the separate strategy, while the dashed one is the fundamental (not necessary obtainable) bound (92).

the cost closely approaches the bound. Consider the largest possible p-dimensional ball inside the simplex $\sum_i |\mu_i| \leqslant \frac{1}{2}$ and then as $f(\boldsymbol{\mu})$ choose the function which minimizes the kinetic energy inside this ball with the boundary condition $f(\boldsymbol{\mu}) = 0$ on the border and outside of the ball. The cost corresponding to this construction, the bound Eq. (92), the values of analytical solution of Eq. (88) for $p = 1, 2$ as well as $\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP}}$ are plotted together in Fig. 3 (details of the calculations may be found in Appendix D). For large p, the total cost corresponding to the described strategy leads to p^3/N^2, and we finally get

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \simeq \frac{cp^3}{N^2}, \quad 0.63 \leqslant c_1 \leqslant 1. \tag{93}$$

When comparing to Eq. (79), we see that in the MM scenario not only the bound is not tight, but it even fails to properly predict the scaling of the cost with the number of parameter p.

To summarize, for the problem of estimation of the spatially distributed single-component magnetic field, with full freedom in choosing both the position and orientation of the atoms, for large p we obtain

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{SEP}} \simeq \frac{p^2}{kn^2}, \quad \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP}} \simeq \frac{\pi^2 p^3}{N^2},$$
$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT}} \simeq \frac{p^2}{kn^2}, \quad \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \simeq \frac{c_1 p^3}{N^2}, \quad 0.63 \leqslant c_1 \leqslant 1. \tag{94}$$

If one wants to compare these results with the previous example (where the positions of the atoms were fixed), one should bear in mind that in the previous example one gate corresponded to p atoms, not one. Therefore, in order to make the comparison fair, one should rewrite Eqs. (70) and (71) in terms of the atoms used $N_a = pN$, $n_a = pn$, which

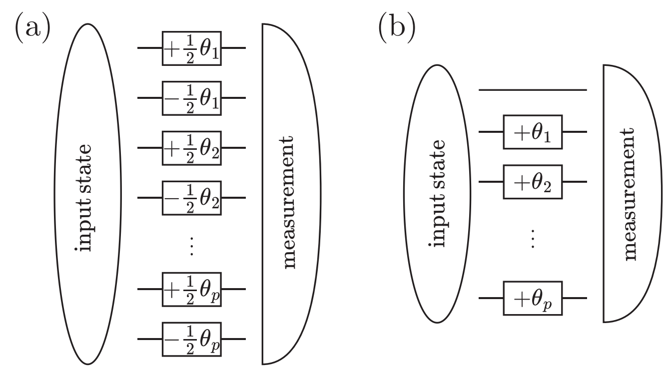

FIG. 4. The problem of estimating spatial distribution of magnetic field sensed by N spin-$\frac{1}{2}$ atoms, with full freedom in choosing their position and orientation may be equivalently seen as the problem of p phases sensing in a $2p$-arms interferometer using N photons (a). This is a slightly modified version of the problem discussed in [61], where p phases were measured in the presence of a single reference arm in $(p+1)$-arms inteferometer (b).

gives

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{SEP+}} \simeq \frac{p^3}{kn_a^2}, \quad \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP+}} \simeq \frac{\pi^2 p^4}{N_a^2}$$
$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT}} \simeq \frac{p^3}{kn_a^2}, \quad \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \simeq \frac{\pi^2 p^3}{N_a^2}. \tag{95}$$

Then it is clear that for the same amount of atoms used, all the costs from Eq. (95) are larger than the corresponding ones from Eq. (94), as in Eq. (95) fewer degrees of freedom are allowed.

Finally, one may notice that the problem of estimating magnetic field in p points by N atoms (with full freedom in choosing their spin and position) is equivalent to a slightly modified problem of multiphase estimation discussed in [61]. Indeed, treating the points in space as arms of an interferometer and atoms as photons, we may think about this problem in terms of a $2p$-arm inferferometer with phase shifts $\pm\frac{1}{2}\theta_i$ in each arm. Therefore, it makes sense to compare Eq. (94) with the analog costs obtained in [61] for the problem of estimating p unknown phase shifts in a $(p+1)$-arm interferometer (where the one arm is the reference arm); both versions are schematically presented in Fig. 4. In the latter case, the single-photon Hilbert space is spanned by $\{|0\rangle, |1\rangle, \dots, |p\rangle\}$ (where $|0\rangle$ corresponds to the reference arm, and $|i\rangle$ to the sensing arms), and the quantum gate is $U_{\boldsymbol{\theta}} = \exp(i\sum_{i=1}^{p}\theta_i |i\rangle\langle i|)$. For large p,

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{SEP}} \simeq \frac{p^2}{kn^2}, \quad \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP}} \simeq \frac{\pi^2 p^3}{N^2},$$
$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT}} \simeq \frac{p^2}{4kn^2}, \quad \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \simeq \frac{c_2 p^3}{N^2}, \quad 1.89 \leqslant c_2 \leqslant 2. \tag{96}$$

Comparing Eq. (94) with Eq. (96), we may notice significant differences. While in the $(p+1)$-arm interferometer, the problem in both the CR and the MM approaches reveals a constant advantage of JNT over SEP strategies of a similar order (around 5), for Eq. (94) no advantage is observed in the CR approach, while the one in the MM approach is even larger (around 10).

This effect, however, may be easily understood after analyzing the structure of the states optimal for measuring parameters separately [Eqs. (76) and (78)]. While in the case studied in [61], in order to estimate any of the parameters one needed to use the same reference arm, in the presently studied model each phase may be measured using two dedicated arms with phase shifts $[-\theta_i/2]$ and $[+\theta_i/2]$. Consequently, the corresponding $|\text{n00n}\rangle$ states [Eq. (76)] are mutually orthogonal: the resources "consumed" by one parameter cannot be used to estimate others. On the other hand, looking at $|\text{SIN}\rangle$ state [Eq. (78)] we see that the greatest weights are attached to the vectors with relatively small differences $m_{i+} - m_{i-} = 0$. As there is no problem in distributing photons (or atoms) in such a way that this difference is small for all i, such a component may be used in estimating all the parameters simultaneously, which is responsible for the significant advantage of the joint strategy in this case.

C. Multicomponent magnetic field estimation

As the last problem, let us discuss the canonical example of estimation of parameters associated with noncommuting generators. More specifically, we focus on the problem of estimating the three components of a magnetic field vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3]^T$ in a given point in space using spin-$\frac{1}{2}$ atoms. Single-atom Hilbert space is therefore simply a qubit space and the corresponding quantum gate reads as

$$U_{\boldsymbol{\theta}} = e^{i\boldsymbol{\theta}\cdot\boldsymbol{\sigma}/2}, \tag{97}$$

where $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \sigma_3]^T$ is the vector of Pauli matrices. Unlike in the previously discussed examples, here the minimal achievable cost depends on the actual values of the parameters $\boldsymbol{\theta}$ [46]. For simplicity, let us focus on estimation around point $\boldsymbol{\theta}_0 = [0, 0, 0]^T$.

Note that for any normalized vector $\boldsymbol{a}$, the operator $\boldsymbol{a}\cdot\boldsymbol{\sigma}$ has the same eigenvalues. Hence, invoking from Eqs. (59) and (60) we see that a rotation in the parameter space cannot improve the precision in a separate protocol; there is no advantage in SEP+.

Let us start by discussing the many-repetition scenario. In a separate strategy, each component may be measured with a proper n00n state $\frac{1}{\sqrt{2}}(|+\rangle_{x,y,z}^{\otimes n} + |-\rangle_{x,y,z}^{\otimes n})$, which leads to

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{CR}}_{\text{SEP}} \simeq 3 \times \frac{1}{k/3} \times \frac{1}{n^2} = \frac{9}{kn^2}. \tag{98}$$

For the joint strategy it turns out that, unlike in the previously discussed examples, application of the adaptive scheme with ancillas allows to beat the performance of the optimal parallel strategy. We will, therefore, discuss these two strategies independently.

The minimal trace of inverse of the QFI achievable for parallel scheme may be found analytically [36] (for $n \geqslant 6$): $\text{Tr}(F^{-1}_{\text{parallel}}) = \frac{9}{n(n+2)}$ [and condition Eq. (23) is satisfied], so

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{CR}}_{\text{JNT parallel}} \simeq \frac{9}{kn(n+2)} \overset{n\gg 1}{\approx} \frac{9}{kn^2}, \tag{99}$$

which for large n is almost the same as for the separate strategy. Therefore, the advantage offered by joint measurement in the parallel strategy disappears with increasing n.

In contrast to the above, it was shown in [45] that for the adaptive ancilla-assisted sequential scheme utilizing as an input state

$$|\psi_{\text{in}}\rangle = \frac{1}{\sqrt{2}}(|+\rangle_z |0\rangle_A + |-\rangle_z |1\rangle_A) \tag{100}$$

(where $|0\rangle_A, |1\rangle_A$ belongs to ancillary system) and acting on it by the gate n times, one may obtain $\text{Tr}(F^{-1}_{\text{adaptive}}) = \frac{3}{n^2}$ [satisfying Eq. (23)], and hence

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{CR}}_{\text{JNT adaptive}} \simeq \frac{3}{kn^2}, \tag{101}$$

which saturates Eq. (53). Therefore, we see that in the many-repetition scenario, the possibility of acting sequentially is crucially needed to take the advantage from the joint estimation approach as it allows to decrease the final variance by a factor 3, compared to the optimal separate or joint parallel strategy.

Consider now the fully optimal usage of N gates. Note that, unlike the previously discussed examples, here the existence of unknown parameters $\theta_{j\neq i}$ may significantly impede estimation of θ_i, as it may be impossible to find an initial state, for which the evolution would depend only on θ_i. As the consequence, we can only write the lower bound:

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{MM}}_{\text{SEP}} \gtrsim 3 \times \frac{\pi^2}{(N/3)^2} = \frac{27\pi^2}{N^2}. \tag{102}$$

Still, it is not obvious how to obtain such a precision while measuring parameters separately.

To calculate the asymptotically optimal cost in joint estimation we use the fact that the problem may be extended to the covariant one, which allows us to use the reasoning from Sec. III C, stating that optimal asymptotical local minimax cost is the same as the global one, obtainable within the parallel scheme. In [55,57], the estimation of a completely unknown element of SU(2) within the parallel strategy was discussed with the covariant cost

$$e(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) = 6 - 2\,\text{Tr}\big(U^{(1)}_{\boldsymbol{\theta}} U^{(1)\dagger}_{\tilde{\boldsymbol{\theta}}}\big) \overset{|\boldsymbol{\theta}|,|\tilde{\boldsymbol{\theta}}|\ll 1}{\approx} 2|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}|^2, \tag{103}$$

where $U^{(1)}_{\boldsymbol{\theta}}$ 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

$$|\psi_{\text{in}}\rangle = \sqrt{\frac{2}{N/2+1}} \sum_{j=0\left(\frac{1}{2}\right)}^{N/2-1} \sin\left(\frac{(j+1)\pi}{J+1}\right) \times \left(\sum_{\alpha=1}^{2j+1} \frac{|j\alpha, m_j = \alpha\rangle}{\sqrt{2j+1}}\right), \tag{104}$$

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). Therefore, including the factor 2 that appears in Eq. (103) we get

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\text{MM}}_{\text{JNT}} \simeq \frac{4\pi^2}{N^2}, \tag{105}$$

which shows a significant advantage over the optimal separate strategy. Based on the discussion in Sec. III C and [82], this result, which is already obtainable within a parallel scheme, cannot be improved by applying any adaptive strategy. This is in stark contrast to the many-repetition scenario, where a parallel joint strategy offers no asymptotical advantage with respect to a separate one, while adaptiveness allows decreasing the cost by a factor of 3. It is also worth noting that the result (105) may be also obtained for parallel strategies with the usage of Fourier analysis, as shown in [[59], Sec. 12].

Analogous reasoning may be performed in a situation when one component of the magnetic field is known to be zero, and only the two remaining components are being estimated. For the separate strategy we get

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{SEP}} \simeq \frac{4}{kn^2}, \quad \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{SEP}} \gtrsim \frac{8\pi^2}{N^2} \tag{106}$$

for the joint CR

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT\,parallel}} \simeq \frac{4}{kn(n+2)}, \quad \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT\,adaptive}} \simeq \frac{2}{kn^2}, \tag{107}$$

while for the joint MM

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \simeq \frac{4\xi^2}{N^2} \approx \frac{2.34\pi^2}{N^2}, \tag{108}$$

where the parallel strategy obtaining above was found in [53,54] and $\xi \approx 2.4048$ is first zero of the Bessel function $J_0(x)$. Equation (108), similarly like Eq. (105), is valid for both parallel and sequential adaptive strategies.

If, on the other hand, the direction of the magnetic field is known and only the length of the magnetic vector is to be estimated, the problem is equivalent to single-phase estimation problem discussed before.

A natural extension of all the above considerations would be to combine all the examples, and consider the most general problem of estimating all the three components of a spatially distributed magnetic field. Based on the analysis performed we expect no improvement in the CR approach and some constant improvement in the MM approach. The strict analysis of this problem, however, is beyond the scope of this paper.

V. CONCLUSIONS

The examples studied demonstrate that, unlike in the single unitary parameter case, in a multiparameter estimation problem there is no simple correspondence between the results obtained within the many-repetition paradigms and the one where all the resources are accumulated in single experimental realization. In the case where the total amount of resources is limited (no matter how large), the analysis based only on the QFI is not sufficient to draw not only quantitative but even qualitative conclusions. The presented examples showed that such an approach tends to overrate the performance of SEP/SEP+ strategies and this opens up the possibility that certain joint estimation metrological strategies may offer a significant advantage, even if this is not apparent in a formalism based on the QFI.

ACKNOWLEDGMENTS

We thank H. M. Wiseman for many fruitful discussions. This work was supported by the National Science Center (Poland) Grant No. 2020/37/B/ST2/02134. W. Górecki was also supported by the Foundation for Polish Science (FNP) via the START scholarship.

APPENDIX A: DIFFERENT APPROACH TO THE IDENTIFICATION OF THE ADVANTAGE OF JOINT ESTIMATION PROTOCOLS

Alternatively to the approaches described in the main text, the issue of the potential gain coming from measuring multiple parameters simultaneously versus measuring them separately may be explored abstracting from the problem of optimal division of resources, but instead by analyzing the so-called probe incompatibility [43,44]. In this approach, the minimal cost achievable in a joint strategy is compared with the one coming from measuring each parameter individually, but with the assumption that in the latter case for each parameter one spends the same amount of resources as in the whole joint strategy (so effectively in the separate strategy p times more resources are consumed).

While in this paper the potential superiority of joint measurement is discussed for a particular cost function (defined by the chosen parametrization for which it is equal to identity), one may instead try to look at the problem as the feature of the channel itself. To do so, in [43] the following quantity was introduced:

$$\mathfrak{J}^* = \max_{\{\mathbf{w}_i\}} \left(\frac{\min_{(n,k)\,\mathrm{protocol}} \mathrm{Tr}(W\Sigma')}{\sum_i \min_{(n,k)\,\mathrm{protocol}} \mathbf{w}_i^T \Sigma' \mathbf{w}_i} \right), \tag{A1}$$

where $W = \sum_i \mathbf{w}_i \mathbf{w}_i^T$. Note that when compared to the original notation [43], we have added a prime sign to Σ' and F', as we reserve the unprimed parametrization for the case where the cost matrix is the identity.

In order to understand this approach, in the context of Fig. 1, let us perform a reparametrization $\boldsymbol{\theta} = A\boldsymbol{\theta}'$ with the transformation matrix $A = [\mathbf{w}_1, \dots, \mathbf{w}_p]^T$. Then, for any fixed set $\{\mathbf{w}_i\}$ we have

$$\frac{\min_{(n,k)\,\mathrm{protocol}} \mathrm{Tr}(W\Sigma')}{\sum_i \min_{(n,k)\,\mathrm{protocol}} \mathbf{w}_i^T \Sigma' \mathbf{w}_i} = \frac{\min_{(n,k)\,\mathrm{protocol}} \mathrm{Tr}(\Sigma)}{\sum_i \min_{(n,k)\,\mathrm{protocol}} \Sigma_{ii}} = \frac{\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT}}}{\sum_i \min_{(n,k)\,\mathrm{protocol}} \Delta^2\tilde{\theta}_i} \tag{A2}$$

so for a fixed $\{\mathbf{w}_i\}$ it corresponds to a situation where Bob alternatively sends nk gates to Alice (nominator) or the same amount of gates directly to each of the minions (denominator). In this sense, the denominator is similar to the SEP strategy (but with omitted problem or resource distribution). Note, however, that the maximization over $\{\mathbf{w}_i\}$ in Eq. (A1) *does not* correspond to SEP+ (as here also the value in the nominator changes while for the reparametrization of A in SEP+ it does not); it should be rather seen as maximization over all possible Bob's initial parametrizations.

APPENDIX B: NECESSITY OF A NONORTHOGONAL TRANSFORMATION IN ORDER TO OBTAIN THE MINIMAL SEP+ COST: EXAMPLE

Here we discuss an exemplary two-parameter estimation problem, for which in order to obtain the minimal cost in

SEP+ protocol we need to apply a nonorthogonal transformation A in the parameter space. We will focus on the many-repetition paradigm and the CR formalism. Consider a unitary channel

$$U_{\boldsymbol{\theta}} = e^{i\boldsymbol{\theta}\cdot\boldsymbol{\Lambda}}, \quad \Lambda_1 = \tfrac{1}{2}\mathrm{diag}(+\alpha, -\alpha, +\beta, -\beta),$$
$$\Lambda_2 = \tfrac{1}{2}\mathrm{diag}(+\beta, -\beta, +\alpha, -\alpha), \tag{B1}$$

where $0 < \beta < \alpha$ and $\mathrm{diag}(\dots)$ is a diagonal matrix acting on the Hilbert space spanned by $|1\rangle, |2\rangle, |3\rangle, |4\rangle$, so

$$\boldsymbol{\theta}\cdot\boldsymbol{\Lambda} = \begin{bmatrix} +\frac{1}{2}(\alpha\theta_1+\beta\theta_2) & 0 & 0 & 0 \\ 0 & -\frac{1}{2}(\alpha\theta_1+\beta\theta_2) & 0 & 0 \\ 0 & 0 & +\frac{1}{2}(\beta\theta_1+\alpha\theta_2) & 0 \\ 0 & 0 & 0 & -\frac{1}{2}(\beta\theta_1+\alpha\theta_2) \end{bmatrix}. \tag{B2}$$

Since the generators mutually commute, the Fisher information matrix for the input $|\psi\rangle$ is given by

$$[F]_{ij} = 4(\langle\psi|\Lambda_i\Lambda_j|\psi\rangle - \langle\psi|\Lambda_i|\psi\rangle\langle\psi|\Lambda_j|\psi\rangle) \tag{B3}$$

and the corresponding CR bound is saturable [as $\mathrm{Im}(\Lambda_1\Lambda_2) = 0$]. By a direct calculation one can see that the cost of joint estimation $\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT}}$ is minimized for the state $|\psi\rangle = \frac{1}{2}(|1\rangle + |2\rangle + |3\rangle + |4\rangle)$ and that for this state

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT}} \simeq \frac{1}{k}\mathrm{Tr}(F^{-1}) = \frac{1}{k}\left(\frac{2}{(\alpha-\beta)^2} + \frac{2}{(\alpha+\beta)^2}\right). \tag{B4}$$

In order to calculate $\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{SEP+}}$, instead of performing a direct optimization given in Eq. (27), we just use the fact that $\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{SEP+}} \geqslant \Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{JNT}}$ and show a particular transformation A for which this bound is saturated. Let us choose as new parameters $\theta_1' = \alpha\theta_1 + \beta\theta_2$, $\theta_2' = \beta\theta_1 + \alpha\theta_2$, so

$$A^{-1} = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix} \Rightarrow A = \frac{1}{\alpha^2-\beta^2}\begin{bmatrix} \alpha & -\beta \\ -\beta & \alpha \end{bmatrix}. \tag{B5}$$

After such a transformation, the differences between extreme eigenvalues of new generators are both equal $\forall_i\, \lambda[A^T\boldsymbol{\Lambda}_i] = 1$, and the new parameters may be effectively measured with states $\frac{1}{\sqrt{2}}(|1\rangle + |2\rangle)$, $\frac{1}{\sqrt{2}}(|3\rangle + |4\rangle)$, which leads to

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{SEP+}} \simeq \frac{1}{k}\left(\sum_{i=1}^{p}\sqrt{[A^2]_{ii}}\right)^2 = \frac{1}{k}\left(\frac{2}{(\alpha-\beta)^2} + \frac{2}{(\alpha+\beta)^2}\right), \tag{B6}$$

which is indeed equal to Eq. (B4). Finally, let us show that this result is not achievable, if one restricted just orthogonal transformations O. Let

$$O^{-1} = [\boldsymbol{o}_1, \quad \boldsymbol{o}_2]^T = \begin{bmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix}. \tag{B7}$$

Then, $[O^TO]_{ii} = 1$ and $[O^{-1}F_i^{-1}O^{-1T}]_{ii} = \boldsymbol{o}_i^T F_i^{-1}\boldsymbol{o}_i$ so

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{CR}}_{\mathrm{SEP+,ortho}} \simeq \frac{1}{k}\min_{\varphi}\left(\sum_{k=1}^{2}\sqrt{\min_{|\psi_k\rangle}\boldsymbol{o}_i^T[F_i]^{-1}\boldsymbol{o}_i}\right)^2. \tag{B8}$$

Both Eqs. (B6) and (B8) are compared in Fig. 5 for different ratios between α and β. One may see that for the ratio around $\frac{1}{2}$ a significant advantage due to application of the nonorthogonal transformation may be observed.

APPENDIX C: RELATION BETWEEN THE OPTIMAL GLOBAL GROUP-INVARIANT COST AND THE LOCAL QUADRATIC MINIMAX COST IN THE LIMIT OF LARGE N

In this Appendix, we formalize the reasoning from Sec. III C.

Theorem. Consider a quantum channel U_g, $g \in G$, which corresponds to a unitary representation of a compact group G in some Hilbert space $U_{g_1}U_{g_2} = U_{g_1g_2}$, and the cost function is invariant with respect to the action of the group $\forall_{g,\tilde{g},h\in G}\mathrm{cost}(hg, h\tilde{g}) = \mathrm{cost}(g, \tilde{g})$. Let $\boldsymbol{\theta} = [\theta_1, \dots, \theta_p]$ be a local parametrization around some $g_0 \in G$. Let $G_\delta \subset G$ be the subset of G containing all $g_{\boldsymbol{\theta}}$ such that $\forall_i\, \theta_i \in [-\delta/2, +\delta/2]$. Then, for the most general adaptive scheme, the local asymptotic minimax cost is the same as the global asymptotic minimax cost:

$$\inf_{\{M_{N,\tilde{g}},\rho_{N,g}\}}\lim_{\delta\to0}\lim_{N\to\infty}N^2\sup_{g\in G_\delta}\int d\tilde{g}\,\mathrm{Tr}(M_{N,\tilde{g}}\rho_{N,g})\mathrm{cost}(g,\tilde{g})$$
$$= \inf_{\{M_{N,\tilde{g}},\rho_{N,g}\}}\lim_{N\to\infty}N^2\sup_{g\in G}\int d\tilde{g}\,\mathrm{Tr}(M_{N,\tilde{g}}\rho_{N,g})\mathrm{cost}(g,\tilde{g}). \tag{C1}$$

Moreover, as obtaining optimal global asymptotic minimax cost was proven not to require adaptiveness [82], the above equation remains valid when restricting the right-hand side to parallel strategies. It also implies that there is no asymptotic advantage in applying adaptive strategy also in the local case.

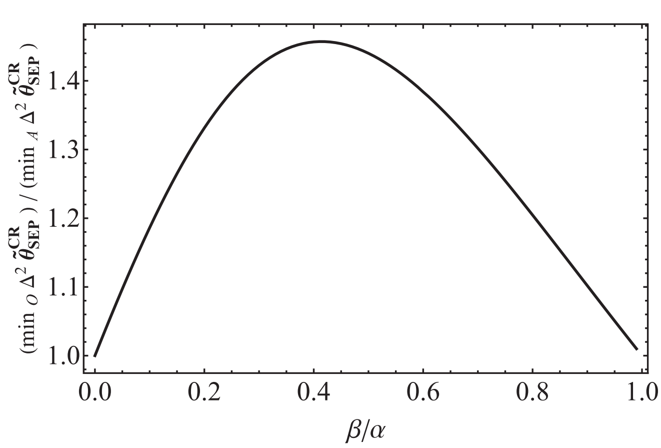

FIG. 5. The ratio between the minimal cost in a separate strategy achievable via optimization over orthogonal reparametrization and a general one.

Proof. Let us introduce the notation for the minimax cost with a finite δ, N:

$$\text{minimax}(G_\delta, N) = \inf_{M_{N,\tilde{g}},\rho_{N,g}} \sup_{g\in G_\delta} \int d\tilde{g}\, \text{Tr}(M_{\tilde{g}}\rho_{N,g})\text{cost}(g,\tilde{g}). \tag{C2}$$

We would like to prove that in the limit of large N, the value of δ has no impact on the final cost. It is clear that

$$\forall_\delta \; \text{minimax}(G_\delta, N) \leqslant \text{minimax}(G, N). \tag{C3}$$

To bound the cost for finite δ from below, we use the following construction [27,61]. Having at our disposal N gates in total we may at first perform $\sqrt{N}$-independent measurements to find an approximated value g_{est}. Due to the central limit theorem, the probability that $g \notin g_{\text{est}}G_\delta$ ($g_{\text{est}}G_\delta$ is the set G_δ shifted by the action of g_{est}) decreases exponentially $p_{\text{err}}(\sqrt{N}) \propto e^{-\sqrt{N}}$. Next, we spend the remaining $N-\sqrt{N}$ gates to perform estimation around the point g_{est}. Since from the point of view of the initial problem of estimating an unknown g with N gates, such a procedure might be suboptimal, we have

$$\begin{aligned}&\text{minimax}(G,N)\\ &\quad\leqslant p_{\text{err}}(\sqrt{N})c_{\max}\\ &\qquad + [1-p_{\text{err}}(\sqrt{N})]\text{minimax}(g_{\text{est}}G_\delta, N-\sqrt{N}),\end{aligned} \tag{C4}$$

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

$$\begin{aligned}\forall_\delta \lim_{N\to\infty} N^2\text{minimax}(G,N) &\leqslant \lim_{N\to\infty} N^2(G_\delta, N-\sqrt{N})\\ &= \lim_{N\to\infty} N^2(G_\delta, N),\end{aligned} \tag{C5}$$

which together with Eq. (C3) gives

$$\forall_\delta \lim_{N\to\infty} N^2\text{minimax}(G_\delta,N) = \lim_{N\to\infty} N^2\text{minimax}(G,N). \tag{C6}$$

Finally, since in general for any functional family $\lim_{x\to x_0}\inf_y F_x(y) \leqslant \inf_y \lim_{x\to x_0} F_x(y)$, the left-hand side of Eq. (C1) may be bounded from below by $\lim_{\delta\to 0}\lim_{N\to\infty} N^2\text{minimax}(G_\delta, N) = \lim_{N\to\infty} N^2\text{minimax}(G,N)$, while the right-hand side is exactly equal to $\lim_{N\to\infty} N^2\text{minimax}(G,N)$. This ends the proof. ■

It is worth to note that if all the elements of the group representation commute $[U_{g_1}, U_{g_2}] = 0$, then in order to implement the strategy (C4) one does not need adaptiveness and the procedure may be performed within the parallel scheme. Indeed, in such a case there is a single state optimal for local measurements around an arbitrary point g_{est}, as rotating the state is equivalent to rotating the measurement in the opposite direction $\text{Tr}(MU_gU_{g_{\text{est}}^{-1}}\rho_0U_{g_{\text{est}}^{-1}}^\dagger U_g^\dagger) = \text{Tr}(U_{g_{\text{est}}^{-1}}^\dagger MU_{g_{\text{est}}^{-1}}U_g\rho_0U_g^\dagger)$, so the knowledge of the value of g_{est} is not needed at the level of state preparation. This is, however, no longer true if $[U_{g_1}, U_{g_2}] \neq 0$. Still, regardless of this suboptimal strategy used in the proof, a fully optimal strategy does not require adaptiveness and may be performed within parallel scheme [82], so theorem remains valid even when restricted to parallel strategies.

Quadratic cost approximation

As shown in [[84], Sec. 9] [[85], Sec. II-D] in local estimation, analyzed within the many-repetition scenario, even if one considers general cost function, in the limit $k\to\infty$ it may be well approximated by the quadratic term $\text{Tr}(W\Sigma) = (\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})^TW(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})$ (where W is the Hessian of the cost function). However, in principle, it is not so clear in the single-repetition approach, as while in the many-repetition scenario, due to the central limit theorem, all probabilities converge to the Gaussian ones (with exponentially decreasing tails), in the single-shot case the strategy minimizing the cost may lead to much slower decreasing tails [75].

More formally, let W be the Hessian of the cost function around g_0, i.e., $W_{ij} = \partial_{\theta_i}\partial_{\theta_j}\text{cost}(g_0, g_{\boldsymbol{\theta}})$. Then for the particular strategy $\{M_{N,\tilde{\boldsymbol{\theta}}}, \rho_{N,\boldsymbol{\theta}}\}$ which is known to lead to the cost $\propto\frac{1}{N^2}$ it is not clear if in the point $\boldsymbol{\theta} = [0,\dots,0]$:

$$\begin{aligned}&\lim_{N\to\infty} N^2\int d\tilde{\boldsymbol{\theta}}\,\text{Tr}(M_{N,\tilde{\boldsymbol{\theta}}}\rho_{N,0})\text{cost}(g_0, g_{\tilde{\boldsymbol{\theta}}})\\ &\quad\overset{?}{=} \lim_{N\to\infty} N^2\int d\tilde{\boldsymbol{\theta}}\,\text{Tr}(M_{N,\tilde{\boldsymbol{\theta}}}\rho_{N,\boldsymbol{\theta}})\sum_{ij}W_{ij}\tilde{\theta}_i\tilde{\theta}_j\end{aligned} \tag{C7}$$

as in principle it may happen that $\text{Tr}(M_{N,\tilde{\boldsymbol{\theta}}}\rho_{N,0})$ decrease like $\propto\frac{1}{N^2|\boldsymbol{\theta}|^2}$ for large $\boldsymbol{\theta}$ [making also higher derivatives of $\text{cost}(g_0, g_{\tilde{\boldsymbol{\theta}}})$ not negligible]. Below we prove that for the strategy minimizing the left-hand side of above, such a tail always decreases fast enough (because if it were otherwise, one could use some small part of gates $\sim\sqrt{N}$ to cut this tail).

Theorem. For the optimal (both adaptive or parallel) global strategy $\{M_{N,\tilde{g}}^{\text{opt}}, \rho_{N,g}^{\text{opt}}\}$, invariant under the group action, for the cost $\text{cost}(g_0, g_{\tilde{\boldsymbol{\theta}}}) = \sum_{ij}W_{ij}\tilde{\theta}_i\tilde{\theta}_j + o(\tilde{\theta}_i\tilde{\theta}_j)$ (where $W > 0$),

$$\begin{aligned}&\lim_{N\to\infty} N^2\int d\tilde{g}\,\text{Tr}\big(M_{N,\tilde{g}}^{\text{opt}}\rho_{N,g_0}^{\text{opt}}\big)\text{cost}(g_0,\tilde{g})\\ &\quad= \lim_{N\to\infty} N^2\int d\tilde{\boldsymbol{\theta}}\,\text{Tr}\big(M_{N,g_{\tilde{\boldsymbol{\theta}}}}^{\text{opt}}\rho_{N,g_0}^{\text{opt}}\big)\sum_{ij}W_{ij}\tilde{\theta}_i\tilde{\theta}_j.\end{aligned} \tag{C8}$$

Proof. For any g we split the mean cost integral into two parts:

$$\begin{aligned}\forall_{g\in G}&\int_G d\tilde{g}\,\text{Tr}\big(M_{N,\tilde{g}}^{\text{opt}}\rho_{N,g}^{\text{opt}}\big)\text{cost}(g,\tilde{g})\\ &= \underbrace{\int_{G_\delta} d\tilde{g}\,\text{Tr}\big(M_{N,\tilde{g}}^{\text{opt}}\rho_{N,g}^{\text{opt}}\big)\text{cost}(g,\tilde{g})}_{C_1(N,\delta)}\\ &\quad+\underbrace{\int_{G\setminus G_\delta} d\tilde{g}\,\text{Tr}\big(M_{N,\tilde{g}}^{\text{opt}}\rho_{N,g}^{\text{opt}}\big)\text{cost}(g,\tilde{g})}_{C_2(N,\delta)}.\end{aligned} \tag{C9}$$

Next, we prove by contradiction that $\forall_{\delta>0}\lim_{N\to\infty} N^2C_2(\delta,N) = 0$.

Assume, that $\lim_{N\to\infty} N^2C_2(\delta,N) > 0$. For a given δ let us choose finite neighborhoods of g_0, namely, G_{Δ_1}, G_{Δ_2}, such that $G_{\Delta_1}\subset G_{\Delta_2}$ and $\forall_{g\in G_{\Delta_1}} G_{\Delta_2}\subset gG_\delta$, satisfying $\max_{g\in G_{\Delta_1}}\text{cost}(g_0,g) < \min_{g\in G_{\Delta_1},\tilde{g}\in G\setminus G_{\Delta_2}}\text{cost}(g,\tilde{g})$ (see left part of Fig. 6).

Next, similarly as in the previous proof, one may at first spend $\sqrt{N}$ gates to find g_{est}, such that the probability that the

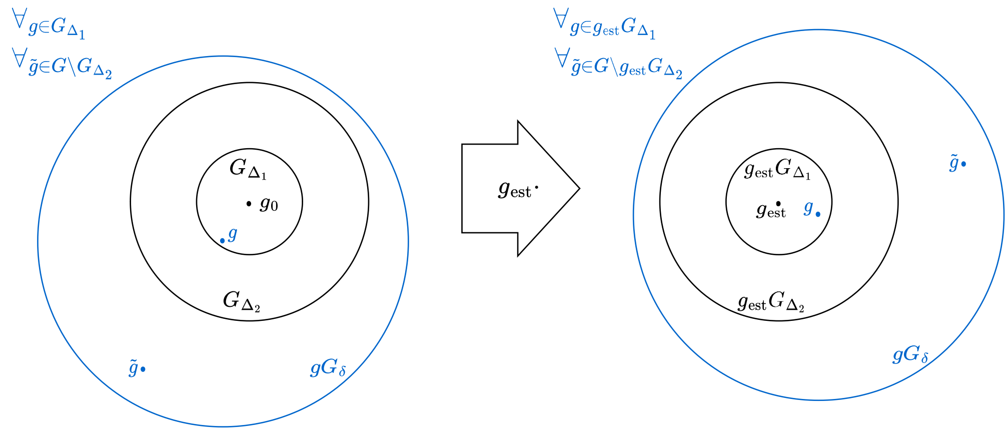

FIG. 6. Graphical illustration of the proof. One spends at first $\sqrt{N}$ gates to find $g_{\rm est}$ such that the probability of finding true value of g outside of $g_{\rm est}G_{\Delta_1}$ is negligible. Then, each time where the second part of the measurement point $\tilde g \notin g_{\rm est}G_{\Delta_2}$, it is more efficient to set $\tilde g = g_{\rm est}$.

true value of g lays outside of $g_{\rm est}G_{\Delta_1}$ decreases exponentially $p_{\rm err}(\sqrt{N}) \propto e^{-\sqrt{N}}$. Then, one uses the remaining $N-\sqrt{N}$ gates to perform the mentioned $M^{\rm opt}_{N-\sqrt{N},\tilde g}\rho^{\rm opt}_{N-\sqrt{N},g}$ with the following correction: each time when $\tilde g$ points outside of $g_{\rm est}G_{\Delta_2}$, one is forced to estimate $\tilde g = g_{\rm est}$. The total cost for such a constructed strategy may be bounded from above:

$$\begin{aligned}&\leqslant p_{\rm err}(\sqrt{N})c_{\max} + [1-p_{\rm err}(\sqrt{N})][C_1(N-\sqrt{N},\delta)\\ &\quad + C_\Delta(N-\sqrt{N},\delta)],\end{aligned} \tag{C10}$$

where $c_{\max} = \max_{g,\tilde g\in G}\operatorname{cost}(g,\tilde g)$ and

$$\begin{aligned}C_\Delta(N-\sqrt{N},\delta) = &\left(\int_{G\backslash G_\delta} d\tilde g\, \operatorname{Tr}\big(M^{\rm opt}_{N-\sqrt{N},\tilde g}\rho^{\rm opt}_{N-\sqrt{N},g}\big)\right)\\ &\max_{g\in G_{\Delta_1}} \operatorname{cost}(g_0,g).\end{aligned} \tag{C11}$$

Since

$$\begin{aligned}C_2(N-\sqrt{N},\delta) \geqslant &\left(\int_{G\backslash G_\delta} d\tilde g\, \operatorname{Tr}\big(M^{\rm opt}_{N-\sqrt{N},\tilde g}\rho^{\rm opt}_{N-\sqrt{N},g}\big)\right)\\ &\min_{g\in G_{\Delta_1},\tilde g\in G\backslash G_{\Delta_2}} \operatorname{cost}(g,\tilde g),\end{aligned} \tag{C12}$$

we have $\lim_{N\to\infty} N^2(C_2(N-\sqrt{N},\delta) - C_\Delta(N-\sqrt{N},\delta)) > 0$. It means that cutting the tail indeed decreases the cost, which leads to a contradiction with the assumption about the optimality of $\{M^{\rm opt}_{N,\tilde g},\rho^{\rm opt}_{N,g}\}$. Now, as from $\mathcal{C}(g_0,g_{\tilde{\boldsymbol\theta}}) = W_{ij}\tilde\theta_i\tilde\theta_j + o(\tilde\theta_i\tilde\theta_j)$ we have $\forall_\epsilon\, \exists_\delta\, \forall_{-\delta/2\leqslant\tilde\theta_i\leqslant+\delta/2}|\mathcal{C}(g_0,g_{\tilde{\boldsymbol\theta}}) - W_{ij}\theta_i\theta_j| \leqslant \epsilon$, the statement is proven. ■

From the above, if $W_{ij} = \delta_{ij}$, then for the optimal strategies the cost is equivalent to $\Delta^2\tilde{\boldsymbol\theta}$ discussed in the main paper. Below we show that this is indeed the case for examples discussed in Sec. IV C. In [55] the authors considered the problem of transmitting a reference frame by sending N spin-$\frac{1}{2}$ atoms, which is equivalent to the estimation of a completely unknown element of SU(2). As the figure of merit they chose the error function given as

$$e(g,\tilde g) = 6 - 2\operatorname{Tr}\big(U^{(1)}_g U^{(1)\dagger}_{\tilde g}\big), \tag{C13}$$

where $U^{(1)}_g$ is the rotation matrix of the spin-1 particle. By a direct calculation

$$\operatorname{Tr}\big(U^{(1)}_{g_0} U^{(1)\dagger}_{g_{\tilde{\boldsymbol\theta}}}\big) = 1 + 2\cos(|\boldsymbol\theta|), \tag{C14}$$

so

$$e(g_0,g_{\tilde{\boldsymbol\theta}}) = 4[1-\cos(|\tilde{\boldsymbol\theta}|)] = 8\sin^2(|\tilde{\boldsymbol\theta}|/2). \tag{C15}$$

Therefore, indeed we have

$$\operatorname{cost}(g_0,g_{\tilde{\boldsymbol\theta}}) = \tfrac{1}{2}e(0,g_{\tilde{\boldsymbol\theta}}) = 4\sin^2(|\tilde{\boldsymbol\theta}|/2) = \tilde{\boldsymbol\theta}^2 + o(\tilde\theta_i\tilde\theta_j). \tag{C16}$$

In [53,54] the authors considered the problem of transmitting a unit vector $\vec n$ by sending N spin-$\frac{1}{2}$ atoms. After introducing parametrization $\vec n = [\cos(\vartheta)\cos(\varphi), \cos(\vartheta)\sin(\varphi), \sin(\vartheta)]$ with $\varphi = \sqrt{\theta_1^2+\theta_2^2}$ and $\vartheta = \arctan(\theta_1/\sqrt{\theta_1^2+\theta_2^2})$, the problem is equivalent to the estimation of the channel $U_{g_{\boldsymbol\theta}} = e^{i(\theta_1\sigma_x/2+\theta_2\sigma_y/2)}$. As the figure of merit they chose fidelity:

$$F(\vec n,\tilde{\vec n}) = (1\pm\vec n\cdot\tilde{\vec n})/2, \tag{C17}$$

for which we have

$$F(\vec n_0,\vec n_{\tilde{\boldsymbol\theta}}) = [1+\cos(|\tilde{\boldsymbol\theta}|)]/2 = 1-\sin^2(|\tilde{\boldsymbol\theta}|/2), \tag{C18}$$

so indeed

$$\begin{aligned}\operatorname{cost}(\vec n_0,\vec n_{\tilde{\boldsymbol\theta}}) &= 4[1-F(\vec n_0,\vec n_{\tilde{\boldsymbol\theta}})] = 4\sin^2(|\tilde{\boldsymbol\theta}|/2)\\ &= \tilde{\boldsymbol\theta}^2 + o(\tilde\theta_i\tilde\theta_j).\end{aligned} \tag{C19}$$

APPENDIX D: DERIVATION OF THE BOUND AND AN EXEMPLARY STATE FOR ESTIMATION OF MAGNETIC FIELD IN *p* POINTS OF SPACE

1. Analytical solutions for *p* = 1, 2

For $p=1$ the simplex $\sum_{i=1}^p|\mu_i| \leqslant 1/2$ is simply the line $\mu_1\in[-1/2,+1/2]$, so the optimal solution is $f(\mu_1) = \sqrt{2}\cos(\pi\mu_1)$ with the cost π^2/N^2. For $p=2$ the simplex $\sum_{i=1}^{p}|\mu_i| \leqslant 1/2$ takes the form of a square with side $\sqrt{2}/2$ rotated by the angle 45° relative to the coordinate axes, which allows for an effective coordinate separation; therefore, the solution is $f(\mu_1,\mu_2) = 2\cos[\sqrt{2}\pi(\mu_1+\mu_2)/\sqrt{2}]\cos[\sqrt{2}\pi(\mu_1-\mu_2)/\sqrt{2}]$ with the corresponding cost $\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} = 2\times(\sqrt{2})^2\pi^2/N^2 = 4\pi^2/N^2$.

2. Derivation of the bound

For the problem (91),

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \geqslant \frac{p}{N^2}\min_g \int_0^{+\infty} d\mu\, g^*(\mu)\left(-\frac{\partial^2}{\partial^2\mu}\right)g(\mu), \quad \text{with} \quad \left.\frac{\partial g}{\partial \mu}\right|_{\mu=0} = 0, \tag{D1}$$

$$\int_0^{+\infty} d\mu |g(\mu)|^2 = 1, \tag{D2}$$

$$\int_0^{+\infty} d\mu |g(\mu)|^2|\mu| = \frac{1}{2p}, \tag{D3}$$

the solution may be found using the standard Lagrange multiplier method

$$-\frac{\partial^2}{\partial\mu^2}g(\mu) + g(\mu)(\lambda_1+\mu\lambda_2) = 0 \Rightarrow g(\mu) \propto \mathrm{Ai}\big(\lambda_2^{1/3}\big(\lambda_1\lambda_2^{-1/3}+\mu\big)\big), \tag{D4}$$

where $\mathrm{Ai}(\cdot)$ is the Airy function of the first kind. The condition $\frac{\partial g}{\partial \mu}|_{\mu=0} = 0$ implies $\lambda_1\lambda_2^{-1/3} = A_0' \approx -1.019$, where A_0' is the first zero of derivative of $\mathrm{Ai}(\cdot)$. From Eqs. (D1) and (D2) we have

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \geqslant \frac{p}{N^2}\frac{\int_0^{+\infty} d\mu\big|\partial_\mu \mathrm{Ai}\big[\lambda_2^{1/3}(A_0'+\mu)\big]\big|^2}{\int_0^{+\infty} d\mu\big|\mathrm{Ai}\big[\lambda_2^{1/3}(A_0'+\mu)\big]\big|^2} = \frac{\int_0^{+\infty} d\mu|\partial_\mu \mathrm{Ai}[(A_0'+\mu)]|^2}{\int_0^{+\infty} d\mu|\mathrm{Ai}[(A_0'+\mu)]|^2}\lambda_2^{2/3}. \tag{D5}$$

To get the value of λ_2 we use Eqs. (D3) and (D2):

$$\frac{1}{2p} = \frac{\int_0^{+\infty} d\mu\big|\mathrm{Ai}\big[\lambda_2^{1/3}(A_0'+\mu)\big]\big|^2\mu}{\int_0^{+\infty} d\mu\big|\mathrm{Ai}\big[\lambda_2^{1/3}(A_0'+\mu)\big]\big|^2} = \frac{\int_0^{+\infty} d\mu|\mathrm{Ai}(A_0'+\mu)|^2\mu}{\int_0^{+\infty} d\mu|\mathrm{Ai}(A_0'+\mu)|^2}\lambda_2^{-1/3}, \tag{D6}$$

so finally

$$\Delta^2\tilde{\boldsymbol{\theta}}^{\mathrm{MM}}_{\mathrm{JNT}} \geqslant \frac{p^3}{N^2}\frac{4\big(\int_0^{+\infty} d\mu|\partial_\mu \mathrm{Ai}[(A_0'+\mu)]|^2\big)\big(\int_0^{+\infty} d\mu|\mathrm{Ai}[(A_0'+\mu)]|^2\mu\big)^2}{\big(\int_0^{+\infty} d\mu|\mathrm{Ai}[(A_0'+\mu)]|^2\big)^3} \approx \frac{0.63p^3}{N^2}. \tag{D7}$$

3. Exemplary state

Finally, let us present a suboptimal, but an explicit, analytical solution of the initial problem (88), which shows a significant advantage compared with the optimal SEP protocol. Namely, we choose the largest possible p-dimensional ball inside the simplex $\sum_i|\mu_i| \leqslant \frac{1}{2}$ and then take as the $f(\boldsymbol{\mu})$ the function which minimizes the kinetic energy inside this ball with a boundary condition $f(\boldsymbol{\mu}) = 0$ on the border and outside of the ball. The Laplacian for spherical coordinated is given as

$$\Delta f = \frac{\partial^2 f}{\partial r^2} + \frac{p-1}{r}\frac{\partial f}{\partial r} + \text{angular part}, \tag{D8}$$

where the exact form of the angular part is irrelevant for the discussion. The corresponding eigenstates are the ones of the form

$$f(r) \propto r^{(2-p)/2}J_{p/2-1}(\sqrt{E}r) \Rightarrow -\Delta f(r) = Ef(r), \tag{D9}$$

where $J_\alpha(\cdot)$ is the Bessel function of the first kind. As the radius of the biggest ball inside the simplex R satisfies

$$R^2 = \sum_{i=1}^{p}(1/2p)^2 = \frac{1}{4p}, \tag{D10}$$

and taking into account the boundary condition $f(R) = 0$ we have

$$\sqrt{E}\frac{1}{2\sqrt{p}} = j_{p/2-1,1} \Rightarrow E = p(2j_{p/2-1,1})^2, \tag{D11}$$

where $j_{p/2-1,1}$ is the zero of the Bessel function $J_{p/2-1}(x)$. Since for large p we have $j_{p/2-1,1} \approx p/2$, we get

$$E \approx p^3, \tag{D12}$$

which yields the joint cost $\frac{p^3}{N^2}$.

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum Metrology, Phys. Rev. Lett. 96, 010401 (2006).

[2] M. G. A. Paris, Quantum estimation for quantum technologies, Int. J. Quantum Inf. 07, 125 (2009).

[3] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photon. **5**, 222 (2011).

[4] G. Toth and I. Apellaniz, Quantum metrology from a quantum information science perspective, J. Phys. A: Math. Theor. **47**, 424006 (2014).

[5] R. Demkowicz-Dobrzanski, M. Jarzyna, and J. Kołodyński, Quantum limits in optical interferometry, in *Progress in Optics*, edited by E. Wolf (Elsevier, Amsterdam, 2015), Vol. 60, pp. 345–435.

[6] R. Schnabel, Squeezed states of light and their applications in laser interferometers, Phys. Rep. **684**, 1 (2017).

[7] C. L. Degen, F. Reinhard, and P. Cappellaro, Quantum sensing, Rev. Mod. Phys. **89**, 035002 (2017).

[8] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Quantum metrology with nonclassical states of atomic ensembles, Rev. Mod. Phys. **90**, 035005 (2018).

[9] S. Pirandola, B. R. Bardhan, T. Gehring, C. Weedbrook, and S. Lloyd, Advances in photonic quantum sensing, Nat. Photon. **12**, 724 (2018).

[10] C. M. Caves, Quantum-mechanical noise in an interferometer, Phys. Rev. D **23**, 1693 (1981).

[11] M. J. Holland and K. Burnett, Interferometric Detection of Optical Phase Shifts at the Heisenberg Limit, Phys. Rev. Lett. **71**, 1355 (1993).

[12] H. Lee, P. Kok, and J. P. Dowling, A quantum rosetta stone for interferometry, J. Mod. Opt. **49**, 2325 (2002).

[13] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen, Spin squeezing and reduced quantum noise in spectroscopy, Phys. Rev. A **46**, R6797 (1992).

[14] K. McKenzie, D. A. Shaddock, D. E. McClelland, B. C. Buchler, and P. K. Lam, Experimental Demonstration of a Squeezing-Enhanced Power-Recycled Michelson Interferometer for Gravitational Wave Detection, Phys. Rev. Lett. **88**, 231102 (2002).

[15] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Optimal frequency measurements with maximally correlated states, Phys. Rev. A **54**, R4649 (1996).

[16] D. Leibfried, M. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. Itano, J. Jost, C. Langer, and D. Wineland, Toward heisenberg-limited spectroscopy with multiparticle entangled states, Science **304**, 1476 (2004).

[17] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: beating the standard quantum limit, Science **306**, 1330 (2004).

[18] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, Improvement of Frequency Standards with Quantum Entanglement, Phys. Rev. Lett. **79**, 3865 (1997).

[19] M. de Burgh and S. D. Bartlett, Quantum methods for clock synchronization: Beating the standard quantum limit without entanglement, Phys. Rev. A **72**, 042301 (2005).

[20] B. Escher, R. de Matos Filho, and L. Davidovich, General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology, Nat. Phys. **7**, 406 (2011).

[21] R. Demkowicz-Dobrzański and L. Maccone, Using Entanglement Against Noise in Quantum Metrology, Phys. Rev. Lett. **113**, 250801 (2014).

[22] W. Dür, M. Skotiniotis, F. Frowis, and B. Kraus, Improved Quantum Metrology Using Quantum Error Correction, Phys. Rev. Lett. **112**, 080801 (2014).

[23] R. Demkowicz-Dobrzański, J. Czajkowski, and P. Sekatski, Adaptive Quantum Metrology under General Markovian Noise, Phys. Rev. X **7**, 041009 (2017).

[24] D. Layden, S. Zhou, P. Cappellaro, and L. Jiang, Ancilla-Free Quantum Error Correction Codes for Quantum Metrology, Phys. Rev. Lett. **122**, 040502 (2019).

[25] S. Zhou, M. Zhang, J. Preskill, and L. Jiang, Achieving the heisenberg limit in quantum metrology using quantum error correction, Nat. Commun. **9**, 78 (2018).

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

[27] M. Hayashi, Comparison between the cramer-rao and the mini-max approaches in quantum channel estimation, Commun. Math. Phys. **304**, 689 (2011).

[28] M. J. W. Hall and H. M. Wiseman, Heisenberg-style bounds for arbitrary estimates of shift parameters including prior information, New J. Phys. **14**, 033040 (2012).

[29] L. Pezzé, Sub-heisenberg phase uncertainties, Phys. Rev. A **88**, 060101(R) (2013).

[30] D. W. Berry, M. Tsang, M. J. W. Hall, and H. M. Wiseman, Quantum Bell-Ziv-Zakai Bounds and Heisenberg Limits for Waveform Estimation, Phys. Rev. X **5**, 031018 (2015).

[31] M. Hayashi, S. Vinjanampathy, and L. C. Kwek, Resolving unattainable cramer–rao bounds for quantum sensors, J. Phys. B: At. Mol. Opt. Phys. **52**, 015503 (2019).

[32] W. Górecki, R. Demkowicz-Dobrzański, H. M. Wiseman, and D. W. Berry, π-corrected Heisenberg Limit, Phys. Rev. Lett. **124**, 030501 (2020).

[33] A. Luis and J. Peřina, Optimum phase-shift estimation and the quantum description of the phase difference, Phys. Rev. A **54**, 4564 (1996).

[34] V. Bužek, R. Derka, and S. Massar, Optimal Quantum Clocks, Phys. Rev. Lett. **82**, 2207 (1999).

[35] D. W. Berry and H. M. Wiseman, Optimal States and Almost Optimal Adaptive Measurements for Quantum Interferometry, Phys. Rev. Lett. **85**, 5098 (2000).

[36] P. Kolenderski and R. Demkowicz-Dobrzanski, Optimal state for keeping reference frames aligned and the platonic solids, Phys. Rev. A **78**, 052333 (2008).

[37] M. Tsang, H. M. Wiseman, and C. M. Caves, Fundamental Quantum Limit to Waveform Estimation, Phys. Rev. Lett. **106**, 090401 (2011).

[38] D. W. Berry, M. J. W. Hall, and H. M. Wiseman, Stochastic Heisenberg Limit: Optimal Estimation of a Fluctuating Phase, Phys. Rev. Lett. **111**, 113601 (2013).

[39] M. Tsang, R. Nair, and X.-M. Lu, Quantum Theory of Superresolution for Two Incoherent Optical Point Sources, Phys. Rev. X **6**, 031033 (2016).

[40] M. G. Genoni, S. Mancini, and A. Serafini, Optimal feedback control of linear quantum systems in the presence of thermal noise, Phys. Rev. A **87**, 042333 (2013).

[41] J. Liu and H. Yuan, Control-enhanced multiparameter quantum estimation, Phys. Rev. A **96**, 042114 (2017).

[42] R. Nichols, P. Liuzzo-Scorpo, P. A. Knott, and G. Adesso, Multiparameter gaussian quantum metrology, Phys. Rev. A **98**, 012114 (2018).

[43] F. Albarelli and R. Demkowicz-Dobrzanski, Probe Incompatibility in Multiparameter Noisy Quantum Metrology, Phys. Rev. X **12**, 011039 (2022).

[44] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański, Compatibility in multiparameter quantum metrology, Phys. Rev. A **94**, 052108 (2016).
[45] H. Yuan, Sequential Feedback Scheme Outperforms the Parallel Scheme for Hamiltonian Parameter Estimation, Phys. Rev. Lett. **117**, 160801 (2016).
[46] T. Baumgratz and A. Datta, Quantum Enhanced Estimation of a Multidimensional Field, Phys. Rev. Lett. **116**, 030801 (2016).
[47] M. Gessner, L. Pezzè, and A. Smerzi, Sensitivity Bounds for Multiparameter Quantum Metrology, Phys. Rev. Lett. **121**, 130503 (2018).
[48] N. Kura and M. Ueda, Finite-error metrological bounds on multiparameter hamiltonian estimation, Phys. Rev. A **97**, 012101 (2018).
[49] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, Distributed Quantum Metrology with Linear Networks and Separable Inputs, Phys. Rev. Lett. **121**, 043604 (2018).
[50] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).
[51] K. Matsumoto, A new approach to the cramér-rao-type bound of the pure-state model, J. Phys. A.: Math. Gen. **35**, 3111 (2002).
[52] R. Demkowicz-Dobrzański, W. Górecki, and M. Guţă, Multiparameter estimation beyond quantum fisher information, J. Phys. A: Math. Theor. **53**, 363001 (2020).
[53] 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 (2000).
[54] 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 (2001).
[55] 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 (2004).
[56] E. Bagan, M. Baig, and R. Muñoz-Tapia, Quantum reverse engineering and reference-frame alignment without nonlocal correlations, Phys. Rev. A **70**, 030301(R) (2004).
[57] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Optimal estimation of group transformations using entanglement, Phys. Rev. A **72**, 042338 (2005).
[58] M. Hayashi, Parallel treatment of estimation of SU(2) and phase estimation, Phys. Lett. A **354**, 183 (2006).
[59] M. Hayashi, Fourier analytic approach to quantum estimation of group action, Commun. Math. Phys. **347**, 3 (2016).
[60] J. Kahn, Fast rate estimation of a unitary operation in SU(d), Phys. Rev. A **75**, 022326 (2007).
[61] W. Górecki and R. Demkowicz-Dobrzański, Multiple-Phase Quantum Interferometry: Real and Apparent Gains of Measuring All the Phases Simultaneously, Phys. Rev. Lett. **128**, 040504 (2022).
[62] P. C. Humphreys, M. Barbieri, A. Datta, and I. A. Walmsley, Quantum Enhanced Multiple Phase Estimation, Phys. Rev. Lett. **111**, 070403 (2013).
[63] A. Z. Goldberg, I. Gianani, M. Barbieri, F. Sciarrino, A. M. Steinberg, and N. Spagnolo, Multiphase estimation without a reference mode, Phys. Rev. A **102**, 022230 (2020).
[64] Ying-Wang, L.-Y. Hu, Z.-M. Zhang, and C.-P. Wei, Simultaneously estimating two phases in three-mode nonlinear interferometer, Opt. Commun. **427**, 250 (2018).
[65] R. Yousefjani, R. Nichols, S. Salimi, and G. Adesso, Estimating phase with a random generator: Strategies and resources in multiparameter quantum metrology, Phys. Rev. A **95**, 062307 (2017).
[66] H. Chen and H. Yuan, Optimal joint estimation of multiple rabi frequencies, Phys. Rev. A **99**, 032122 (2019).
[67] L. B. Ho, H. Hakoshima, Y. Matsuzaki, M. Matsuzaki, and Y. Kondo, Multiparameter quantum estimation under dephasing noise, Phys. Rev. A **102**, 022602 (2020).
[68] S. Pirandola and C. Lupo, Ultimate Precision of Adaptive Noise Estimation, Phys. Rev. Lett. **118**, 100502 (2017).
[69] J. Suzuki, Nuisance parameter problem in quantum estimation theory: Tradeoff relation and qubit examples, J. Phys. A: Math. Theor. **53**, 264001 (2020).
[70] J. Suzuki, Y. Yang, and M. Hayashi, Quantum state estimation with nuisance parameters, J. Phys. A: Math. Theor. **53**, 453001 (2020).
[71] J. Häjek, Local asymptotic minimax and admissibility in estimation, in *Theory of Statistics* (University of California Press, Oakland, 1972), pp. 175–194.
[72] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, The elusive heisenberg limit in quantum-enhanced metrology, Nat. Commun. **3**, 1063 (2012).
[73] S. Zhou and L. Jiang, Asymptotic theory of quantum channel estimation, PRX Quantum **2**, 010343 (2021).
[74] A. Chiruvelli and H. Lee, Parity measurements in quantum optical metrology, J. Mod. Opt. **58**, 945 (2011).
[75] H. Imai and M. Hayashi, Fourier analytic approach to phase estimation in quantum systems, New J. Phys. **11**, 043034 (2009).
[76] T. Kaftal and R. Demkowicz-Dobrzański, Usefulness of an enhanced kitaev phase-estimation algorithm in quantum metrology and computation, Phys. Rev. A **90**, 062313 (2014).
[77] 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 (2009).
[78] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde, Entanglement-free heisenberg-limited phase estimation, Nature (London) **450**, 393 (2007).
[79] B. Higgins, D. Berry, S. Bartlett, M. Mitchell, H. Wiseman, and G. Pryde, Demonstrating heisenberg-limited unambiguous phase estimation without adaptive measurements, New J. Phys. **11**, 073023 (2009).
[80] H. Imai and A. Fujiwara, Geometry of optimal estimation scheme for SU(d) channels, J. Phys. A: Math. Theor. **40**, 4391 (2007).
[81] J. E. Cohen, S. Friedland, T. Kato, and F. P. Kelly, Eigenvalue inequalities for products of matrix exponentials, Lin. Alg. Applicat. **45**, 55 (1982).
[82] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Memory Effects in Quantum Channel Discrimination, Phys. Rev. Lett. **101**, 180501 (2008).
[83] S. Rey and P. Choquard, The calogero - sutherland - moser (CSM) model as a quantum billiard, Eur. J. Phys. **18**, 94 (1997).
[84] Y. Yang, G. Chiribella, and M. Hayashi, Attaining the ultimate precision limit in quantum state estimation, Commun. Math. Phys. **368**, 223 (2019).
[85] M. Hayashi and K. Matsumoto, Asymptotic performance of optimal state estimation in qubit system, J. Math. Phys. **49**, 102101 (2008).