Determining which quantum measurement performs better for state estimation

Jaroslav Řeháček, Yong Siah Teo, and Zdeněk Hradil

Department of Optics, Palacký University, 17 listopadu 12, 77146 Olomouc, Czech Republic

We introduce an operational and statistically meaningful measure, the quantum tomographic transfer function, that possesses important physical invariance properties for judging whether a given informationally complete quantum measurement performs better tomographically in quantum-state estimation relative to other informationally complete measurements. This function is independent of the unknown true state of the quantum source and is directly related to the average optimal tomographic accuracy of an unbiased state estimator for the measurement in the limit of many sampling events. For the experimentally appealing minimally complete measurements, the transfer function is an extremely simple formula. We also give an explicit expression for this transfer function in terms of an ordered expansion that is readily computable and illustrate its usage with numerical simulations and its consistency with some known results.

I. INTRODUCTION

Quantum-state estimation constitutes a broad class of tools catered to the verification and diagnostics of states for quantum systems, a necessary step in accessing the validity of quantum protocols. Thus far, much work has been devoted to the development of statistical methods for quantum-state reconstruction [1–3], to the investigation of the tomographic accuracies of quantum-state estimators [4–6], to improving the performance of state reconstruction under difficult circumstances (such as large Hilbert-space dimensions, noisy and ill-calibrated detections, etc.) with various approaches [7–10], and to the assignment of statistical measures that quantify the error and reliability of state estimators [11,12], and so forth.

This article focuses primarily on the comparison of different quantum measurements available for state estimation, which is yet another important aspect in this field. In particular, a value of a certain statistically meaningful quantity is assigned to every quantum measurement that serves as a gauge on their tomographic performance. The ascription of a measurement [13–15] with a meaningful number that quantifies its performance is not a recent proposal. In optics, one can describe the performance of an optical system with the optical transfer function [16], which incorporates the details of signal propagation from every single component in the optical system to the image plane. In quantum-state estimation, it is also useful to introduce such a function to every quantum measurement. Further assessment of tomographic accuracies can be subsequently carried out by other means (see, for instance, Ref. [17]).

II. BACKGROUND

In quantum-state estimation, the unknown quantum state ρ of a source is reconstructed from a set of measurement data supplied by the quantum measurement apparatus, which is mathematically represented by a probability-operator measure (POM) consisting of M positive operators Π_j: $\sum_j \Pi_j = 1$. For a D-dimensional Hilbert space, ρ can always be written as a linear combination of D^2 Hermitian, trace-orthonormal basis operators, where one operator is a multiple of the identity $1/\sqrt{D}$, and the rest of the $D^2 - 1$ operators Ω_j are traceless.

This operator basis conveniently incorporates the unit-trace constraint of the state, $\text{tr}\{\rho\} = 1$, and organizes all the relevant $D^2 - 1$ independent state parameters to be estimated. The correct probabilities $p_j = \text{tr}\{\rho \, \Pi_j\}$ for the source and apparatus are related to the state ρ according to Born's rule, which translates to a linear matrix equation $\mathbf{p}' = \mathcal{C}\mathbf{t}$, where $\mathbf{t}$ is a column of $D^2 - 1$ independent coefficients $\text{tr}\{\rho \, \Omega_j\}$, $\mathbf{p}'$ has components $p_j - \text{tr}\{\Pi_j\}/D$, and the central object of our discussion, the $M \times (D^2 - 1)$ measurement matrix $\mathcal{C}$ with entries $\mathcal{C}_{jk} = \text{tr}\{\Pi_j \Omega_k\}$, provides complete information about the apparatus. Throughout this discussion, we shall consider only informationally complete POMs ($M \geqslant D^2$), that is, POMs that uniquely characterize ρ. A POM is minimally complete if it contains $M = D^2$ outcomes that are all linearly independent.

There exists a well-known concept of the condition number (ratio of the largest to the smallest singular value of the measurement matrix) that quantifies the error stability for the experiment. The preferred measurement is usually one that maximizes this stability or, in other words, has the smallest condition number κ out of the conceivable choice. It can be easily shown that this number is simply inadequate to assess the tomographic accuracies of state estimators, and such a preferred choice is not always the one that maximizes the tomographic accuracy. The pitfalls in comparing different quantum measurements solely by studying the condition numbers of $\widetilde{\mathcal{C}}$ are discussed in Sec. I.

The appropriate quantity to consider for tomographic-accuracy estimation is the trace of the inverse of the scaled Fisher information matrix $\boldsymbol{F}(\rho) = \mathcal{C}^\mathsf{T} \mathcal{P}^{-1} \mathcal{C}$ for the usual multinomial detection statistics (Sec. III) as the meaningful quantity in determining the performance of a POM, with $\mathcal{P} = \text{diag}(p_1, p_2, \ldots, p_M)$. This quantity is well motivated. First, if the unknown state ρ is full rank, which is always the case in realistic experiments, the matrix trace $\text{Sp}\{\boldsymbol{F}(\rho)^{-1}\}$ gives the scaled (over sampling events) optimal tomographic accuracy for all unbiased state estimators $\widehat{\rho}$ in the limit of large sampling events. This accuracy is given as the Hilbert-Schmidt distance between ρ and $\widehat{\rho}$, which is a natural distance for quantum states and has been investigated in, for instance, Refs. [4–6]. Second, the scaled Fisher matrix $\boldsymbol{F}(\rho)$ possesses an important physical invariance property—it is invariant under channel duplication (Sec. I). Given a POM with M

outcomes, no additional information is obtained if any part of this POM is precisely duplicated.

To define a measure that is independent of the unknown state, one can perform an average over the state space to obtain an averaged performance for a given POM. Since the state of the source is unknown, and nearly pure states are typically the quantum states of interest in many quantum protocols, this average may be taken in an uninformative way [18], in the sense that the Haar prior for pure states is used for the average.

III. MAIN RESULTS

A. A power series

Unfortunately, state averages on $\boldsymbol{F}(\rho)^{-1}$ are generally difficult to calculate directly, and this difficulty begins with the fact that $(\boldsymbol{\mathcal{C}}^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{\mathcal{C}})^{-1} = \boldsymbol{\mathcal{C}}^{-}\boldsymbol{\mathcal{P}}(\boldsymbol{\mathcal{C}}^{\mathrm{T}})^{-}$ *does not hold in general* for generalized inverses $\boldsymbol{\mathcal{C}}^{-}$ of $\boldsymbol{\mathcal{C}}$. Instead, the average is carried out termwise on an expansion of the right-hand side of the identity

$$\mathrm{Sp}\{\boldsymbol{F}(\rho)^{-1}\} = \mathrm{Sp}\{\overline{\boldsymbol{F}}^{-1}\} + \mathrm{Sp}\left\{\frac{\boldsymbol{\mathcal{X}}\Delta\boldsymbol{\mathcal{P}}}{1 - \boldsymbol{\mathcal{Y}}\Delta\boldsymbol{\mathcal{P}}}\right\} \quad (1)$$

in powers of $\Delta\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{P}} - \overline{\boldsymbol{\mathcal{P}}}$. Here, $\overline{\boldsymbol{\mathcal{P}}}$ is the diagonal matrix of probabilities $\overline{p}_j = \mathrm{tr}\{\Pi_j\}/D$ for the maximally mixed state, $\overline{\boldsymbol{F}} = \boldsymbol{F}(1/D)$, $\boldsymbol{\mathcal{X}} = \overline{\boldsymbol{\mathcal{P}}}^{-1}\boldsymbol{\mathcal{C}}\overline{\boldsymbol{F}}^{-2}\boldsymbol{\mathcal{C}}^{\mathrm{T}}\overline{\boldsymbol{\mathcal{P}}}^{-1}$, and $\boldsymbol{\mathcal{Y}} = \overline{\boldsymbol{\mathcal{P}}}^{-1}\boldsymbol{\mathcal{C}}\overline{\boldsymbol{F}}^{-1}\boldsymbol{\mathcal{C}}^{\mathrm{T}}\overline{\boldsymbol{\mathcal{P}}}^{-1} - \overline{\boldsymbol{\mathcal{P}}}^{-1}$. The positive matrices $\boldsymbol{\mathcal{X}}$ and $-\boldsymbol{\mathcal{Y}}$ have several interesting properties. First of all, they are mutually conjugate orthogonal: $\boldsymbol{\mathcal{X}}\overline{\boldsymbol{\mathcal{P}}}\boldsymbol{\mathcal{Y}} = 0 = \boldsymbol{\mathcal{Y}}\overline{\boldsymbol{\mathcal{P}}}\boldsymbol{\mathcal{X}}$. The matrix $-\overline{\boldsymbol{\mathcal{P}}}$ turns out to be the generalized inverse of $\boldsymbol{\mathcal{Y}}$: $\boldsymbol{\mathcal{Y}}\overline{\boldsymbol{\mathcal{P}}}\boldsymbol{\mathcal{Y}} = -\boldsymbol{\mathcal{Y}}$. The matrix $\boldsymbol{\mathcal{Y}}$ also belongs to the kernel of $\boldsymbol{\mathcal{C}}$: $\boldsymbol{\mathcal{C}}^{\mathrm{T}}\boldsymbol{\mathcal{Y}} = 0 = \boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{C}}$.

A power series for $\mathrm{Sp}\{\boldsymbol{F}(\rho)^{-1}\}$ exists only when the eigenvalues of $\|\boldsymbol{\mathcal{Y}}\Delta\boldsymbol{\mathcal{P}}\|$ are sufficiently small. It is possible to take advantage of the scaling transformation $\boldsymbol{F}(\alpha\rho) = \boldsymbol{F}(\rho)/\alpha$ to obtain convergence series for any ρ by setting $\alpha < \alpha_0 = 1/(\|Y\|_2 \max_j\{\mathrm{tr}\{\Pi_j\}\})$ (see also Sec. III). After taking the Haar average of this series, we define the tomographic measure for any POM as

qTTF($\{\Pi_j\}$)

$\equiv \overline{\mathrm{Sp}\{\boldsymbol{F}(\rho)^{-1}\}}$

$= \underbrace{\mathrm{Sp}\{\overline{\boldsymbol{F}}^{-1}\}}_{\text{zeroth order}} + \underbrace{\frac{\alpha}{D(D+1)}\sum_{j_1,j_2=1}^{M}\boldsymbol{\mathcal{X}}_{j_2 j_1}\boldsymbol{\mathcal{Y}}_{j_1 j_2}\boldsymbol{\mathcal{G}}_{j_1 j_2}^{(2)}}_{\text{second order}} + \cdots, \quad (2)$

where $\boldsymbol{\mathcal{G}}_{j_1 j_2 \ldots j_n}^{(n)} = \mathrm{tr}\{\Pi_{j_1}\Pi_{j_2}\ldots\Pi_{j_n}\}$ are elements of high-order Gram matrices, which is the main result of this article (Sec. IV). As this function keeps track of all connections related to the quantum measurement with the average optimal tomographic accuracy of an unbiased $\widehat{\rho}$, we shall name it the *quantum tomographic transfer function* (qTTF) for convenience, in a similar spirit as the optical transfer function in classical optics. Terminologies and acronyms aside, one should remember that this function was defined on a solid conceptual framework in statistics.

One can now understand the fallacies arising from solely studying the condition number of $\boldsymbol{\mathcal{C}}$. In its real singular-value decomposition, $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{O}}'^{\mathrm{T}}$, with $\boldsymbol{\mathcal{O}}^{\mathrm{T}}\boldsymbol{\mathcal{O}} =$

$1 = \boldsymbol{\mathcal{O}}'\boldsymbol{\mathcal{O}}'^{\mathrm{T}} = \boldsymbol{\mathcal{O}}'^{\mathrm{T}}\boldsymbol{\mathcal{O}}'$, $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{O}}'$ of dimensions $(D^2 - 1) \times (D^2 - 1)$ and $\boldsymbol{\mathcal{O}}$ of dimensions $M \times (D^2 - 1)$, the quantity $\mathrm{Sp}\{\boldsymbol{F}(\rho)^{-1}\} = \mathrm{Sp}\{(\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{O}}^{\mathrm{T}}\boldsymbol{\mathcal{P}}^{-1}\boldsymbol{\mathcal{O}}\boldsymbol{\mathcal{S}})^{-1}\}$ depends on *both* the diagonal matrix $\boldsymbol{\mathcal{S}}$ of singular values, as well as the orthogonal matrix $\boldsymbol{\mathcal{O}}$. While $\boldsymbol{\mathcal{S}}$ loosely describes the distribution of tomographic weights given to the measurement outcomes, the matrix $\boldsymbol{\mathcal{O}}$ contains information about the mutual relationships among these outcomes. The complete information about the optimal tomographic accuracy of $\widehat{\rho}$ that is carried by these two mathematical objects is distributed throughout the series in Eq. (E3), with part of this information manifested in the higher order Gram matrices $\boldsymbol{\mathcal{G}}^{(n)}$. Leaving any of these objects out of the description renders any POM comparison potentially illegitimate. For instance, it is a simple matter to find two POMs such that the one with the bigger κ value (weakly conditioned) gives the lower qTTF value (greater optimal tomographic accuracy).

The higher-order terms in Eq. (E3) determine the difference between the averaged quantity and the quantity evaluated at the maximally mixed state. The formula shows that this difference can only be accounted for by Gram matrices of *all* orders. The comprehensive role of these Gram matrices is now clear—they contain all crucial mutual relationships among the measurement outcomes that enter the tomographic error propagation to $\widehat{\rho}$, a characteristic that is analogous to the mechanism behind the optical transfer function, which encompasses all relative phase information of the signal through the individual components of an optical system.

B. Closed-form formulas for typical measurements

There exists an extremely simple formula for the qTTF when evaluating minimally complete POMs. For POMs of this type, since $\boldsymbol{\mathcal{Y}}$ resides in the kernel of $\boldsymbol{\mathcal{C}}$, the rank of $\boldsymbol{\mathcal{Y}}$ must therefore be one, for $\boldsymbol{\mathcal{C}}$ now has a $(D^2 - 1)$-dimensional row space. Since $\sum_j \boldsymbol{\mathcal{C}}_{jk} = 0$, $\boldsymbol{\mathcal{Y}}$ must therefore be a negative matrix with entries all equal to minus one in the computational basis: $\boldsymbol{\mathcal{Y}}_{j,k} = -1$. Then, we have $\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{Y}} = -\boldsymbol{\mathcal{Y}}$ for any probability matrix $\boldsymbol{\mathcal{P}}$ so that the second term in the right-hand side of Eq. (1) has a finite series expansion consisting of only the two lowest-order terms in $\Delta\boldsymbol{\mathcal{P}}$. Thus, Eq. (E3) with $\alpha = 1$ greatly simplifies to

$$\mathrm{qTTF}(\{\Pi_j\}_{\mathrm{MIN}}) = \mathrm{Sp}\{\overline{\boldsymbol{F}}_{\mathrm{MIN}}^{-1}\} - 1 + \frac{1}{D} \quad (3)$$

for any minimally complete POM $\{\Pi_j\}_{\mathrm{MIN}}$. For overcomplete POMs made up of $D + 1$ bases (minimally complete bases) of D rank-one outcomes [$M = D(D + 1)$], the qTTF is also completely described by only the first two terms of Eq. (E3) with $\alpha = 1$, as it can be shown that $\boldsymbol{\mathcal{Y}}$ is a rank-$(D + 1)$ projector that takes a block-diagonal form, with $D + 1$ blocks each of dimensions $D \times D$ and all matrix elements equal to $-(D + 1)$ for every block, so that again $\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{P}}\boldsymbol{\mathcal{Y}} = -\boldsymbol{\mathcal{Y}}$ and

$$\mathrm{qTTF}(\{\Pi_j\}_{\mathrm{MIN\,BASES}}) = \frac{\mathrm{Sp}\{(\boldsymbol{\mathcal{C}}^{\mathrm{T}}\boldsymbol{\mathcal{C}})^{-1}\}}{(D+1)^2}. \quad (4)$$

The formulas presented in this article can be verified to be consistent with known results for highly symmetric measurements. To this end, we note that the trace of the scaled Fisher matrix $\mathrm{Sp}\{\overline{\boldsymbol{F}}\} \leqslant D(D - 1)$ evaluated at $\boldsymbol{\mathcal{P}} = \overline{\boldsymbol{\mathcal{P}}}$ is bounded

from above by $D(D-1)$ for *any* informationally complete POM. With the help of the inequality $\mathrm{Sp}\{\boldsymbol{\mathcal{A}}\}\mathrm{Sp}\{\boldsymbol{\mathcal{A}}^{-1}\} \geqslant (\dim\{\boldsymbol{\mathcal{A}}\})^2$, which is saturated when $\boldsymbol{\mathcal{A}}$ is a multiple of the identity, we find that the zeroth term in Eq. (E3) is bounded from below inasmuch as

$$\mathrm{Sp}\{\overline{\boldsymbol{F}}^{-1}\} \geqslant \frac{(D+1)(D^2-1)}{D}, \qquad (5)$$

which again holds for any informationally complete POM. For minimally complete POMs, the equality in Eq. (5) is attained for symmetric informationally complete (SIC) POMs, where $\overline{\boldsymbol{F}} = \overline{\boldsymbol{F}}_{\mathrm{MIN}} = \overline{\boldsymbol{F}}_{\mathrm{SIC}} = D/(D+1)$ is a multiple of the (D^2-1)-dimensional identity. We thus obtain the well-known lower bound for the qTTF evaluated with SIC POMs of dimension D to be

$$\mathrm{qTTF}(\{\Pi_j\}_{\mathrm{SIC}}) = D^2 + D - 2 \qquad (6)$$

for minimally complete POMs, as reported in Refs. [4–6]. When D is a prime power, it can be shown that any complete set of mutually unbiased bases (MUB) will also saturate Eq. (5), and since the inequality $\mathrm{Sp}\{\boldsymbol{\mathcal{A}}\} \leqslant a_{\mathrm{max}}\dim\{\boldsymbol{\mathcal{A}}\}$ with respect to the largest eigenvalue a_{max} of $\boldsymbol{\mathcal{A}}$ is saturated when $\boldsymbol{\mathcal{A}}$ is a multiple of the identity, the right-hand side of Eq. (4) is correspondingly minimized to

$$\mathrm{qTTF}(\{\Pi_j\}_{\mathrm{MUB}}) = D^2 - 1 < \mathrm{qTTF}(\{\Pi_j\}_{\mathrm{SIC}}) \qquad (7)$$

over all minimally complete bases [6].

C. General procedures to handle the power series

For any other kind of overcomplete POMs, there is no simple closed-form expression for the series of qTTF in Eq. (E3). Although it is straightforward to obtain higher order terms as desired, the computation becomes exponentially exhaustive as either M or D increases. We shall now highlight the viable procedures for numerically computing the first few computable terms of the qTTF for different regimes of M and D. If both $M > D^2$ and D are not too large, then the qTTF of any POM can be found by computing higher-order corrections in Eq. (E3) for $\alpha \approx \alpha_0$. If D is large, so that the computation of high-order terms starts to become expensive, and yet M is not too far from D^2, it turns out that taking the sum of the zeroth- and second-order terms for $\alpha = 1$ gives good approximation to the qTTF. The largest relative error—the ratio of the difference between this approximation and the actual qTTF to the actual qTTF—as M approaches infinity is given by $D/(D+2)$ in Sec. E. One may also take $\alpha \approx \alpha_0$ and attempt to approximate the series with various models, but numerical experience indicates that terminating the series up to the second-order correction for $\alpha = 1$ yields more accurate approximations in this regime of M and D. If both M and D are large, then performing a Monte Carlo calculation by averaging $\mathrm{Sp}\{\boldsymbol{F}(\rho)^{-1}\}$ over a Haar ensemble of pure states is the most economical way of computing the qTTF.

D. Figures and results

We supply two figures to illustrate the validity of qTTF and its calculation procedures. Figure 1 shows, for various D, the halved relative error of the second-order approximation of qTTF (aqTTF) compared with the correct qTTF value
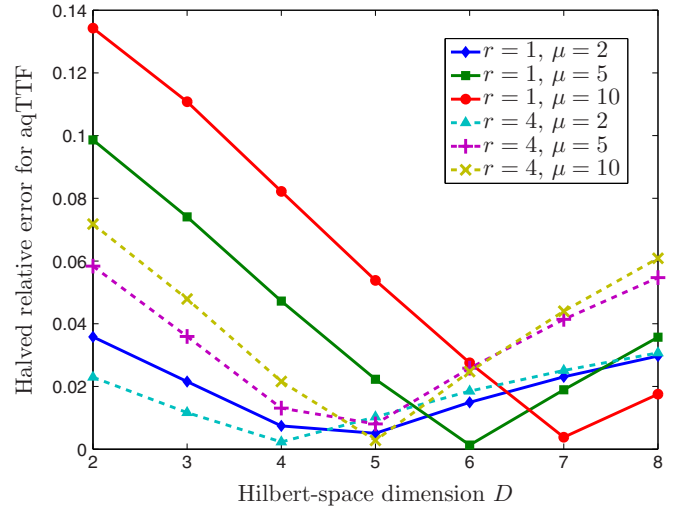

FIG. 1. (Color online) A plot showing the halved relative error of the aqTTF compared with the correct qTTF value. An average over 500 random rank-one ($r = 1$) and full-rank ($r = 4$) POMs is carried out separately to compute the data points for different μ and D. The full-rank POMs are generated by taking a convex sum of rank-one POM outcomes and the maximally mixed state with small admixtures ($\Pi_j = \mathcal{N}_j\{|\psi_j\rangle\langle\psi_j| + 0.05/D\}$) with proper POM normalization $\mathcal{N}_j$. To perform Monte Carlo computations of the qTTF, a Haar set of 500 random pure states is used. For values of μ for which the number of POM outcomes $M = \mu D^2$ is reasonable, such as the ones shown in the plot, the halved relative error decreases with increasing dimension D. The plots indicate that the performance of the second-order correction improves significantly for the slightly mixed POM.

(averaged over many random POMs) obtained by averaging $\mathrm{Sp}\{\boldsymbol{F}(\rho)^{-1}\}$ over a set of random pure states distributed according to the Haar measure [19]. The one-half factor originates from error propagation to $\widehat{\rho}$. We generate random POMs for averaging that mimic those used in real experimental scenarios. A set of M complex positive operators $B_j = A_j^\dagger A_j / \mathrm{tr}\{A_j^\dagger A_j\}$ is first generated, with the $\mathrm{rank}\{\Pi_j\} \times D$ matrices A_j having complex entries distributed according to the standard Gaussian distribution. The resulting POM is obtained with its outcomes given by $\Pi_j = (\sum_j B_j)^{-1/2} B_j (\sum_j B_j)^{-1/2}$ that sum to the identity. The set of random outcomes generated this way tend to have similar traces and this models the typical measurement outcomes employed in an experiment, where the slight variation in the traces originate from systematic instrumental errors and losses that result in nonunit detection efficiencies. Figure 1 tells us that for moderate M, the second-order approximation for Eq. (E3) indeed works rather well even for large D. For extremely large values of M, the halved relative errors approach the limiting value $D/[2(D+2)]$ for rank-one POMs. Witnessing this limit is, however, a rather impractical feat in any experiment. Moreover, in a realistic situation, the POM outcomes designed may have small amounts of white noise, and the aqTTF is significantly more accurate. As the amount of white noise increases, the zeroth-order term gets increasingly more accurate and all other correction terms vanish since $\Delta\mathcal{P} \approx \mathbf{0}$.

Figure 2 compares the values of four quantities: the condition number κ for $\widetilde{\boldsymbol{\mathcal{C}}}'$, the qTTF, the aqTTF, and the scaled
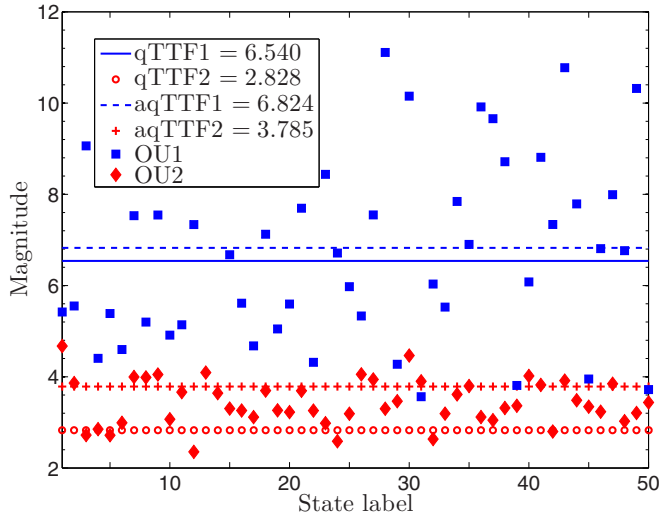
FIG. 2. (Color online) Comparison of different tomographic quantifiers for two chosen qubit ($D = 2$) overcomplete POMs of rank one, with POM 1 associated with $\kappa_1 = 3.437$ and POM 2 associated with $\kappa_2 = 8.119$. A Haar set of 500 random pure states are used for computing the respective qTTFs for both POMs. For both POMs, the averaged Hilbert-Schmidt distance between $\widehat{\rho}$ and ρ is evaluated with 50 random states, each of which is a pure state that is slightly mixed with the maximally mixed state, so that the purity is fixed at ≈ 0.990. It is clear that the average tomographic performance of the OU estimator over all the random states lies near the qTTF and aqTTF for each POM, of course, which tells us that POM 2 performs much better than POM 1 for nearly pure states, whereas an incorrect conclusion would otherwise have been drawn by naively comparing κ_1 and $\kappa_2 > \kappa_1$. ML estimators (not shown) were also computed, and their average performance also lie close to the respective qTTFs and aqTTFs, as they should.

mean squared-error (MSE) of both the optimal-unbiased (OU) and maximum-likelihood (ML) reconstruction schemes [20]. Comparisons are made between a fixed pair of POMs for each of the different D values. This figure gives counterexamples that confirm, once and for all, that the qTTF is the appropriate quantity for estimating the tomographic accuracy for $\widehat{\rho}$, not κ.

IV. CONCLUSIONS

To conclude, we have emphasized the importance of meaningful statistical quantification of a quantum measurement with the quantum tomographic transfer function, which is based on the Haar average of the trace of the inverse Fisher matrix that is equivalent to the optimal tomographic accuracy for unbiased state estimators, as long as the unknown state of the quantum source is not rank deficient, which is the case in any experiment. This transfer function can be thought of as a quantum analog of the optical transfer function for optical systems. We gave an explicit expression for this transfer function as a series [Eq. (E3)] that is readily computable, and provided numerical evidence for the validity of the transfer function in terms of tomographic accuracy estimation of measurements. This function possesses physical invariance properties that are crucial in properly judging the quality of the measurement. Typically, one can take the second-order

approximation of the transfer function as a good approximation as long as the Hilbert-space dimension and the number of outcomes are not too large, and in cases where the known measurement outcomes are full rank due to slight perturbations by white noise, this approximation improves in accuracy. Despite the fact that Monte Carlo computation of the qTTF applies for all quantum measurements, the closed-form formulas [Eqs. (3) and (4)] provide a better understanding about the structural relationships between quantum measurements and their performance.

APPENDIX A: THE CONDITION NUMBER

There are proposals [13–15] that exploit the condition number for judging the quality of a given quantum measurement. We shall briefly demonstrate the shortcomings in deciding which quantum measurement performs better by studying the condition number κ of $\widetilde{\mathcal{C}}$, the $M \times D^2$ measurement matrix constructed from D^2 trace-orthonormal Hermitian basis operators Γ_j. First, there can exist two POMs with exactly the same condition number that perform quite differently in tomography. A perfect example would be the comparison in tomographic performance between SIC POMs ($M = 2^2 = 4$) and MUB for $D = 2$. Upon adopting the set of basis operators

$$\Gamma_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Gamma_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\Gamma_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \Gamma_4 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{(A1)}$$

the respective $\widetilde{\mathcal{C}}$ matrices for the two measurements can be shown to have precisely the same condition number $\kappa = 1.7321$. However, it is well known that the MUB outperform the SIC POMs in terms of lowering the mean squared error between a quantum-state estimator and the true state that does not lie on the boundary [4–6], which is the typical experimental situation. Thus, the condition number fails to distinguish between two such measurements.

Second, apart from this insensitivity, the condition number of $\widetilde{\mathcal{C}}$ has a serious flaw as a tomographic measure—it is not invariant under channel duplication. To illustrate this, we use the basis operators in Eq. (A1) and first write down

$$\widetilde{\mathcal{C}} \cong \begin{pmatrix} 0.3536 & -0.2041 & 0.2041 & 0.2041 \\ 0.3536 & 0.2041 & 0.2041 & -0.2041 \\ 0.3536 & 0.2041 & -0.2041 & 0.2041 \\ 0.3536 & -0.2041 & -0.2041 & -0.2041 \end{pmatrix} \quad \text{(A2)}$$

for the qubit SIC POM with a set of basis operators, whose singular values are $\{0.7071, 0.4082, 0.4082, 0.4082\}$, such that $\kappa = 1.7321$. Next, the fourth outcome of this POM (last row of $\widetilde{\mathcal{C}}$), say, is duplicated with equal proportions, so that the

resulting 5×4 matrix

$$
\widetilde{\mathcal{C}}' \widehat{=} \begin{pmatrix}
0.3536 & -0.2041 & 0.2041 & 0.2041 \\
0.3536 & 0.2041 & 0.2041 & -0.2041 \\
0.3536 & 0.2041 & -0.2041 & 0.2041 \\
0.1768 & -0.1021 & -0.1021 & -0.1021 \\
0.1768 & -0.1021 & -0.1021 & -0.1021
\end{pmatrix} \quad \text{(A3)}
$$

now has singular values $\{0.6700, 0.4082, 0.4082, 0.3047\}$, with $\kappa = 2.1988$. Physically, duplicating any channel(s) cannot affect the Fisher information gain. The condition number κ, and other functions of the singular values of $\widetilde{\mathcal{C}}$ alone for that matter, however, would report completely nonsensical results.

This little excursion tells us an important fact: The condition number cannot be the full story in terms of gauging tomographic accuracies corresponding to different quantum measurements. While this number alone appears to be consistent with the comparison in tomographic accuracy for the three examples given in Ref. [13], for instance, one can find other examples showing that the sole consideration of the condition number is inadequate and can even give contradicting conclusions for overcomplete measurements. Nevertheless, the articles [13] and [14] accurately point out that there are other crucial factors in determining the quality of a quantum-state estimation protocol.

A word of caution: The articles [13–15] compare $M \times D^2$ measurement matrices, which we denote by $\widetilde{\mathcal{C}}$, for the usual linear-inversion (LIN) tomography schemes, rather than comparing the $M \times (D^2 - 1)$ $\mathcal{C}$ matrices discussed here. We point out that the former approach of comparison is incompatible with unit-trace LIN estimators, and this shall be elaborated in Appendix B.

APPENDIX B: LINEAR-INVERSION TOMOGRAPHY

The usual linear-inversion tomography schemes involve optimizing some figures of merit of interest. A very common approach is the minimization of the squared error $S = |\mathbf{f}' - \mathcal{C}\mathbf{t}|^2$, where $\mathbf{f}'$ is a column of components $f_j - \text{tr}\{\Pi_j\}/D$ and f_j are the frequency data for the POM outcomes obtained from the experiment. It is well known that $\mathbf{t} = \mathcal{C}^-\mathbf{f}$ is the solution to the least-squares problem for this quantity, where $\mathcal{C}^-$ is the Moore-Penrose pseudoinverse of $\mathcal{C}$ such that $\mathcal{C}^-\mathcal{C} = 1$, so that the state estimator $\widehat{\rho} = 1/D + \sum_j t_j \Omega_j$. Therefore, it is meaningful to directly compare $\mathcal{C}$ of different POMs.

In the other two proposals, the measurement matrices $\widetilde{\mathcal{C}}$ of dimensions $M \times D^2$ are compared instead. The corresponding least-squares problem involves minimizing the same quantity S introduced in the previous paragraph. It is easy to see that the Moore-Penrose pseudoinverses of these matrices do not guarantee unit-trace state estimators, and that further enforcing the unit-trace constraint will result in the new estimators being nonlinear in $\widetilde{\mathcal{C}}^-$. The least-squares estimators can in fact be shown to be

$$
\widehat{\rho} = \sum_{k=1}^{D^2} [\widetilde{\mathcal{C}}^- + \chi (\widetilde{\mathcal{C}}^\mathsf{T}\widetilde{\mathcal{C}})^{-1} (\mathbf{e} \ \mathbf{e} \ \dots \ \mathbf{e})]_k \Gamma_k,
$$

$$
\chi = \frac{1 - \sqrt{D}\, \mathbf{e}^\mathsf{T} \widetilde{\mathcal{C}}^- \mathbf{f}}{\sqrt{D}\, \mathbf{e}^\mathsf{T} (\widetilde{\mathcal{C}}^\mathsf{T}\widetilde{\mathcal{C}})^{-1} \mathbf{e}}, \quad \text{(B1)}
$$

where Γ_k form a complete set of D^2 Hermitian trace-orthonormal basis operators and $\mathbf{e}^\mathsf{T} \widehat{=} (1,0, \dots, 0)$. The linear-inversion estimators are thus complicated functions of these measurement matrices, and direct comparisons of $\widetilde{\mathcal{C}}$ do not compatibly correspond to the assessment of the quality of these estimators.

The straightforward reason is that the unit-trace constraint is never taken into account automatically in minimizing S. Rather, this constraint is to be additionally fulfilled, for instance, by optimizing only the relevant $D^2 - 1$ independent state parameters that are associated to traceless basis operators, as was done in the main article, or by some other method.

APPENDIX C: FISHER INFORMATION MATRIX FOR MULTINOMIAL STATISTICS

For multinomial statistics of a *fixed* sequence of data, the likelihood is given by

$$
\mathcal{L}(\{n_j\}; \rho) = \prod_k p_k^{n_k}. \quad \text{(C1)}
$$

In terms of the column $\mathbf{t}$ of coefficients t_j for the state $\rho = 1/D + \sum_j t_j \Omega_j$, one can evaluate the term

$$
\frac{\partial}{\partial \mathbf{t}} \log \mathcal{L}(\{n_j\}; \rho) = \sum_k n_k \frac{\partial}{\partial \mathbf{t}} \log p_k
$$

$$
= \sum_k \frac{n_k}{p_k} \frac{\partial p_k}{\partial \mathbf{t}}. \quad \text{(C2)}
$$

The well-known scaled Fisher information dyadic for ρ can then be found to be

$$
\begin{aligned}
\boldsymbol{F}(\rho) &= \frac{1}{N} \overline{\left(\sum_k \frac{n_k}{p_k} \frac{\partial p_k}{\partial \mathbf{t}} \right) \left(\sum_l \frac{n_l}{p_l} \frac{\partial p_l}{\partial \mathbf{t}} \right)} \\
&= \frac{1}{N} \sum_{k,l} \frac{\overline{n_k n_l}}{p_k p_l} \frac{\partial p_k}{\partial \mathbf{t}} \frac{\partial p_l}{\partial \mathbf{t}} \\
&= \frac{N^2 - N}{N} \underbrace{\sum_{k,l} \frac{\partial p_k}{\partial \mathbf{t}} \frac{\partial p_l}{\partial \mathbf{t}}}_{=0} + \sum_l \frac{1}{p_l} \frac{\partial p_l}{\partial \mathbf{t}} \frac{\partial p_l}{\partial \mathbf{t}} \\
&= \sum_l \frac{1}{p_l} \frac{\partial p_l}{\partial \mathbf{t}} \frac{\partial p_l}{\partial \mathbf{t}}. \quad \text{(C3)}
\end{aligned}
$$

Since

$$
p_l = \text{tr}\{\rho \Pi_l\} = \frac{\text{tr}\{\Pi_j\}}{D} + \sum_j t_j \text{tr}\{\Omega_j \Pi_l\},
$$

$$
\frac{\partial p_l}{\partial \mathbf{t}} = \text{tr}\{\boldsymbol{\Omega} \Pi_l\}, \quad \text{(C4)}
$$

we find that

$$
\boldsymbol{F}(\rho) = \sum_l \text{tr}\{\boldsymbol{\Omega} \Pi_l\} \frac{1}{p_l} \text{tr}\{\Pi_l \boldsymbol{\Omega}\} = \mathcal{C}^\mathsf{T} \mathcal{P}^{-1} \mathcal{C}. \quad \text{(C5)}
$$

APPENDIX D: CONVERGENCE OF POWER SERIES

There exists a power series for the left-hand side of

$$\mathrm{Sp}\{\boldsymbol{F}(\rho)^{-1}\} = \frac{1}{\alpha}\left[\mathrm{Sp}\{\overline{\boldsymbol{F}}^{-1}\} + \mathrm{Sp}\left\{\frac{\boldsymbol{\mathcal{X}}(\alpha\boldsymbol{\mathcal{P}}-\overline{\boldsymbol{\mathcal{P}}})}{\boldsymbol{1}-\boldsymbol{\mathcal{Y}}(\alpha\boldsymbol{\mathcal{P}}-\overline{\boldsymbol{\mathcal{P}}})}\right\}\right], \tag{D1}$$

which is just a scaling transformation mentioned in the main text, as long as the largest eigenvalue of $\boldsymbol{\mathcal{Y}}(\alpha\boldsymbol{\mathcal{P}}-\overline{\boldsymbol{\mathcal{P}}})$, denoted by $\sigma_{\max}\{\boldsymbol{\mathcal{Y}}(\alpha\boldsymbol{\mathcal{P}}-\overline{\boldsymbol{\mathcal{P}}})\}$, is small. By using the inequalities

$$\sigma_{\max}\{\boldsymbol{\mathcal{A}}-\boldsymbol{\mathcal{B}}\} \leqslant \max\{\sigma_{\max}\{\boldsymbol{\mathcal{A}}\},\sigma_{\max}\{\boldsymbol{\mathcal{B}}\}\},$$
$$\sigma_{\max}\{\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}\} \leqslant \sigma_{\max}\{\boldsymbol{\mathcal{A}}\}\sigma_{\max}\{\boldsymbol{\mathcal{B}}\} \tag{D2}$$

for any positive matrices $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$, it can be shown that $\alpha < 1/(\|Y\|_2 \max_j\{\mathrm{tr}\{\Pi_j\}\})$ is a sufficient condition for this existence.

APPENDIX E: POWER SERIES

By noting that the Haar average

$$\overline{(|\ \rangle\langle\ |)^{\otimes n}} = \frac{S_n}{\mathrm{tr}\{S_n\}} \tag{E1}$$

over all pure statistical operators of a given Hilbert-space dimension D is related to the permutation projector S_n on the n-fold D-dimensional symmetric subspace, it is possible to derive the series

$$\mathrm{qTTF}(\{\Pi_j\}) = \underbrace{\mathrm{Sp}\{\overline{\boldsymbol{F}}^{-1}\}}_{\text{zeroth order}} + \underbrace{\alpha F_2}_{\text{second order}} + \underbrace{\alpha^2(F_3 - F_2) + \alpha F_2}_{\text{third order}} + \underbrace{\alpha^3(F_4 - 2F_3 + F_2) + 2\alpha^2(F_3 - F_2) + \alpha F_2}_{\text{fourth order}} + \cdots, \tag{E2}$$

where

$$F_2 = \frac{1}{D(D+1)}\sum_{j_1,j_2=1}^{M}\boldsymbol{\mathcal{X}}_{j_2 j_1}\boldsymbol{\mathcal{Y}}_{j_1 j_2}\boldsymbol{\mathcal{G}}^{(2)}_{j_1 j_2},$$

$$F_3 = \frac{2}{D(D+1)(D+2)}\left[\sum_{j_1,j_2=1}^{M}\boldsymbol{\mathcal{X}}_{j_2 j_1}\boldsymbol{\mathcal{Y}}_{j_1 j_2}\boldsymbol{\mathcal{G}}^{(2)}_{j_1 j_2} + \sum_{j_1,j_2,j_3=1}^{M}\boldsymbol{\mathcal{X}}_{j_3 j_1}\boldsymbol{\mathcal{Y}}_{j_1 j_2}\boldsymbol{\mathcal{Y}}_{j_2 j_3}\mathrm{Re}\{\boldsymbol{\mathcal{G}}^{(3)}_{j_1 j_2 j_3}\}\right],$$

$$F_4 = \frac{1}{D(D+1)(D+2)(D+3)}\left[(6-D)\sum_{j_1,j_2=1}^{M}\boldsymbol{\mathcal{X}}_{j_2 j_1}\boldsymbol{\mathcal{Y}}_{j_1 j_2}\boldsymbol{\mathcal{G}}^{(2)}_{j_1 j_2} + 12\sum_{j_1,j_2,j_3=1}^{M}\boldsymbol{\mathcal{X}}_{j_3 j_1}\boldsymbol{\mathcal{Y}}_{j_1 j_2}\boldsymbol{\mathcal{Y}}_{j_2 j_3}\mathrm{Re}\{\boldsymbol{\mathcal{G}}^{(3)}_{j_1 j_2 j_3}\}\right.$$
$$\left. + \sum_{j_1,j_2,j_3,j_4=1}^{M}\boldsymbol{\mathcal{X}}_{j_4 j_1}\boldsymbol{\mathcal{Y}}_{j_1 j_2}\boldsymbol{\mathcal{Y}}_{j_2 j_3}\boldsymbol{\mathcal{Y}}_{j_3 j_4}\left(2\,\mathrm{Re}\{\boldsymbol{\mathcal{G}}^{(4)}_{j_1 j_2 j_3 j_4} + \boldsymbol{\mathcal{G}}^{(4)}_{j_1 j_2 j_4 j_3} + \boldsymbol{\mathcal{G}}^{(4)}_{j_1 j_3 j_2 j_4}\} + \boldsymbol{\mathcal{G}}^{(2)}_{j_1 j_2}\boldsymbol{\mathcal{G}}^{(2)}_{j_3 j_4} + \boldsymbol{\mathcal{G}}^{(2)}_{j_1 j_3}\boldsymbol{\mathcal{G}}^{(2)}_{j_2 j_4} + \boldsymbol{\mathcal{G}}^{(2)}_{j_1 j_4}\boldsymbol{\mathcal{G}}^{(2)}_{j_2 j_3}\right)\right]. \tag{E3}$$

APPENDIX F: SECOND-ORDER CORRECTION TERM

For any nonadaptive quantum-state tomography scheme using a fixed POM of M outcomes, the limiting tomographic performance as M approaches infinity is defined by that of the covariant measurement for a given Hilbert-space dimension D [6], whose outcomes themselves form a Haar ensemble of pure states. Therefore, the relative error between the approximate qTTF, obtained via the second-order correction, and the actual qTTF increases to the upper bound defined by this measurement. To calculate this upper bound, it is enough to argue that the zeroth-order term gives precisely the minimum value $(D+1)(D^2-1)/D$ and show that the difference between the approximate qTTF and the actual value is $4(D^2-1)/(D+2)$. Since the qTTF for the covariant measurement is equal to $2(D-1)$, the relative error is given by $D/(D+2)$.

[1] M. Paris and J. Řeháček, *Quantum State Estimation* (Springer, Berlin, 2004); J. Řeháček, Z. Hradil, E. Knill, and A. I. Lvovsky, Phys. Rev. A **75**, 042108 (2007).

[2] A. R. Rossi and M. G. A. Paris, Eur. Phys. J. D **32**, 223 (2005); Y. S. Teo, H. Zhu, B.-G. Englert, J. Řeháček, and Z. Hradil, Phys. Rev. Lett. **107**, 020404 (2011); Y. S. Teo, B.-G. Englert, J. Řeháček, and Z. Hradil, Phys. Rev. A **84**, 062125 (2011).

[3] R. Schack, T. A. Brun, and C. M. Caves, Phys. Rev. A **64**, 014305 (2001); R. Blume-Kohout and P. Hayden, arXiv:quant-ph/0603116; R. Blume-Kohout, New J. Phys. **12**, 043034 (2010); V. S. Shchesnovich, arXiv:1403.5158.

[4] A. J. Scott, J. Phys. A **39**, 13507 (2006).

[5] H. Zhu and B.-G. Englert, Phys. Rev. A **84**, 022327 (2011).

[6] H. Zhu, Phys. Rev. A **90**, 012115 (2014).

[7] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, New J. Phys. **14**, 095022 (2012); T. Baumgratz, D. Gross, M. Cramer, and M. B. Plenio, Phys. Rev. Lett. **111**, 020401 (2013); C. Schwemmer, G. Toth, A. Niggebaum, T. Moroder, D. Gross, O. Gühne, and H. Weinfurter, *ibid.* **113**, 040503 (2014).

[8] J. Řeháček, D. Mogilevtsev, and Z. Hradil, Phys. Rev. Lett. **105**, 010402 (2010); D. Mogilevtsev, A. Ignatenko, A. Maloshtan, B. Stoklasa, J. Řeháček, and Z. Hradil, New J. Phys. **15**, 025038 (2013); L. Moťka, B. Stoklasa, J. Řeháček, Z. Hradil, V. Karásek, D. Mogilevtsev, G. Harder, C. Silberhorn, and L. L. Sánchez-Soto, Phys. Rev. A **89**, 054102 (2014).

[9] F. Huszár and N. M. T. Houlsby, Phys. Rev. A **85**, 052120 (2012); K. S. Kravtsov, S. S. Straupe, I. V. Radchenko, N. M. T. Houlsby,

F. Huszar, and S. P. Kulik, *ibid.* **87**, 062122 (2013); D. H. Mahler, L. A. Rozema, A. Darabi, C. Ferrie, R. Blume-Kohout, and A. M. Steinberg, Phys. Rev. Lett. **111**, 183601 (2013).

[10] D. Mogilevtsev, J. Řeháček, and Z. Hradil, New J. Phys. **14**, 095001 (2012); S. S. Straupe, D. P. Ivanov, A. A. Kalinkin, I. B. Bobrov, S. P. Kulik, and D. Mogilevtsev, Phys. Rev. A **87**, 042109 (2013); Y. S. Teo, J. Řeháček, and Z. Hradil, *ibid.* **88**, 022111 (2013).

[11] R. Blume-Kohout, arXiv:1202.5270; M. Christandl and R. Renner, Phys. Rev. Lett. **109**, 120403 (2012); **109**, 159903(E) (2012).

[12] J. Shang, H. K. Ng, A. Sehrawat, X. Li, and B.-G. Englert, New J. Phys. **15**, 123026 (2013).

[13] Y. I. Bogdanov, G. Brida, M. Genovese, S. P. Kulik, E. V. Moreva, and A. P. Shurupov, Phys. Rev. Lett. **105**, 010404 (2010).

[14] Y. I. Bogdanov, G. Brida, I. D. Bukeev, M. Genovese, K. S. Kravtsov, S. P. Kulik, E. V. Moreva, A. A. Soloviev, and A. P. Shurupov, Phys. Rev. A **84**, 042108 (2011).

[15] A. Miranowicz, K. Bartkiewicz, J. Peřina Jr., M. Koashi, N. Imoto, and F. Nori, Phys. Rev. A **90**, 062123 (2014).

[16] J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts and Company, Englewood, Colorado, 2004).

[17] M. Bellini, A. S. Coelho, S. N. Filippov, V. I. Man'ko, and A. Zavatta, Phys. Rev. A **85**, 052129 (2012).

[18] K. R. W. Jones, Ann. Phys. **207**, 140 (1991); Phys. Rev. A **50**, 3682 (1994).

[19] K. Życzkowski and H.-J. Sommers, J. Phys. A **34**, 7111 (2001).

[20] Linear-inversion estimators do not minimize the Hilbert-Schmidt MSE measure. The authors refer the reader to Eqs. (6) and (7) in Ref. [6] for the formulas that lead to OU state estimators with respect to this measure.