

Optimal state estimation of controllable quantum dynamical systems

Raj Chakrabarti*

School of Chemical Engineering, Purdue University, West Lafayette, Indiana 47907, USA

Anisha Ghosh†

Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

(Received 25 March 2011; published 7 March 2012)

We present system-theoretic quantum state reconstruction methods that minimize estimation error by combining optimal quantum control with asymptotically efficient estimation. Introducing the notion of optimal observability of a quantum dynamical system—a concept that does not exist in classical control theory—we formulate and solve the Pareto optimal control problem of maximizing state estimation accuracy while minimizing the expenditure of available control and measurement resources. Necessary and sufficient conditions for optimal observability, based on the quantum optimal observability Gramian, are presented. We examine the finite sample efficiency of the estimation methodology for two- and three-level systems using ideal and noisy control fields, and demonstrate the advantages of state reconstruction schemes based on optimal observability theory for experimentally realistic sample sizes. These results indicate that the optimal control of quantum measurement bases can be used to minimize state reconstruction errors by fully exploiting the information geometry of quantum states.

DOI: [10.1103/PhysRevA.85.032305](https://doi.org/10.1103/PhysRevA.85.032305)

PACS number(s): 03.67.–a, 05.30.–d, 02.50.–r, 03.65.Wj

I. INTRODUCTION

Perhaps the most fundamental problem in quantum statistical inference (QSI) is the reconstruction of the density matrix of a quantum system based on a finite number of quantum observations [1–3]. Due to the rapidly growing interest in quantum computation and quantum control [4–7], the ability to retrieve the maximum amount of information about a quantum state based on the smallest number of measurements is a subject of paramount importance. The accuracy of all derivative forms of QSI, including process estimation, is ultimately determined by that of the underlying state estimation.

Frequentist methods for quantum state estimation—which deliver distributional results on estimator accuracy in the limit of an infinite number of measurements—can be formally subdivided into two categories. The first, tomographic inversion [8–12], is the least computationally expensive, and most popular technique. With suitable parametrizations of the density matrix, there is no need to numerically solve an optimization problem, and the parameter estimates can be obtained by solving a system of linear equations. Due to their simplicity they are commonly applied in state reconstruction experiments [13], and are especially popular in the context of process identification [14–16]. However, tomographic inversion cannot enforce the constraints on the density matrix during estimation, and hence the estimates produced are often not physically meaningful. The second class consists of frequentist techniques of inference based on a likelihood function, the most notable of which is the maximum likelihood (ML) estimation [17–21]. This class of methods avoids the problems associated with tomography and is asymptotically more efficient. Banaszek *et al.* [17] and Rehacek *et al.* [20] introduced

numerical algorithms for quantum ML. Recent studies [21] have considered the quantification of state estimation errors within quantum ML, and experimental demonstrations of ML quantum state estimation have also been reported [22].

Frequentist inference of the state of a quantum system based on measurements at a fixed time requires a complete observation level—corresponding to $N^2 - 1$ distinguishable measurement outcomes—in order to estimate all parameters. For systems on high-dimensional Hilbert spaces (e.g., molecules), constructing a complete set of observables through distinct measurements can be difficult or impossible due to limitations on experimentally measurable physical quantities. Bayesian estimation [23–27], which is based on updating a prior plausibility distribution about the parameters based on observed data, is an alternative to frequentist estimation that is capable of state reconstruction despite incomplete observation levels. Bayesian methods are computationally intensive since they require the use of numerical (e.g., Monte Carlo) integration techniques [27] in order to provide parameter estimates and standard errors. The review of Buzek *et al.* [12] surveys techniques for Bayesian state estimation and compares them to frequentist methods.

While estimation methods based on prior plausibility distributions can surmount the problem of incomplete measurements, the accuracy of the resulting state estimates is substantially compromised due to the lack of a complete observation level. However, systems engineering theory provides methods—based on the notion of *observability*—for exploiting the dynamical evolution of a system to completely reconstruct a state despite limitations on the number of available measurement types and in the absence of a prior plausibility distribution. Observability [28,29] is concerned with whether, given a set of observable operators and a dynamical system, the initial state of the system can be obtained by measurements on the system at different times under the influence of control. These systems-theoretic approaches to state reconstruction reported thus far are direct extensions

*rchakra@purdue.edu

†anishagh@andrew.cmu.edu

of classical bilinear theory [30]. The notion of optimality of estimation or control does not enter.

Optimality of quantum state estimation, by contrast, depends on the properties of quantum measurement. Due to the noncommutativity of the quantum probability space, there exist optimal measurement strategies [3,31,32] that deliver more information about the state for an equal number of measurements. There is an extensive and sophisticated literature on optimal quantum estimation that prescribe the choice of measurements that yield the most efficient parameter estimates [33–36], provided that an asymptotically efficient estimator, such as ML, is used.¹ In this paper, we show how quantum control theory [4,6,7,38] can be combined with analytical results from optimal quantum estimation theory to extract maximal information from quantum states using minimal measurement and control resources. We introduce methods for improving the accuracy of quantum state reconstruction methods by combining optimal quantum control, quantum estimation, and observability theory. We incorporate the properties of quantum measurement into observability theory in order to identify necessary and sufficient conditions for optimal observability—the ability to extract maximal information about the state of a controlled quantum dynamical system. We introduce the quantum optimal observability Gramian matrix, as well as other analytical tests, as a means of checking for optimal observability. These concepts have no analog in classical observability theory. We thus provide *optimal system-theoretic state reconstruction schemes*. Since the required control and measurement resources are reduced to the theoretical minimum while the information extracted from the state is maximized, the problem is one of quantum Pareto optimal control [39]. These methods provide prescriptions for how experimentalists can reconstruct quantum states, for systems whose Hamiltonians are known, using measurements of easily accessible observable quantities, such as the energy. This is achieved through the use of flexible electromagnetic field resources, rather than by expanding the set of experimentally observable quantities, as has been the focus of prior art. The application of systems engineering techniques to state reconstruction has been limited by the fact that their efficiency is lower than that of the latest quantum state estimation methods. The present methods display efficiencies that exceed those of standard systems engineering schemes, by leveraging the latest knowledge regarding optimal quantum state estimation.

Optimal state estimation can be achieved by applying controls (e.g., electromagnetic fields) that drive the system dynamical propagator to measurement bases that maximize the extracted Fisher information about the state. As will be shown below, this requires N controlled time evolutions. If a non-degenerate observable is measured, complete state estimation can be achieved with just one type of measurement (for example, repeated measurements of the energy).² The needed controlled

evolutions can be obtained via the application of optimal gate control theory [40,41]. We show that among nonadaptive measurement schemes (i.e. those where measurements do not depend on the outcome of previous measurements), the proposed strategy is asymptotically optimal from the perspective of state reconstruction error for a fixed set of measurement resources.

Among frequentist estimation techniques, maximum likelihood is usually preferred on the basis of its asymptotic properties, e.g., the ML estimator is asymptotically efficient in the sense that its asymptotic variance achieves the Cramér-Rao lower bound for consistent estimators [42]. However, like any frequentist estimator, it delivers distributional results for the estimators of the parameters of interest asymptotically. To interrogate the laboratory feasibility of the proposed scheme, it is necessary to account for the possibility of noise in dynamical parameters as well as the finite sample performance of asymptotically efficient estimators.

We therefore investigate optimal system-theoretic state reconstruction from the following perspectives:

- (i) Given a set of available control and measurement resources, can one globally minimize state reconstruction errors?
- (ii) For optimally controlled state estimation, how do the finite-sample standard errors, and hence the associated 95% confidence intervals, compare to the corresponding asymptotically predicted ones?
- (iii) How robust are the optimal state estimators to noise in the control fields?

Extensive simulation results are provided that demonstrate that the strategy is nearly optimal and robust to noise in small samples.

The paper is organized as follows. Section II details the characteristics of ML estimators of the quantum density matrix and the choice of measurements that asymptotically maximize the information extracted from the state. In Sec. III, we present necessary and sufficient conditions for optimal observability of quantum dynamical systems, compare them to the standard conditions for observability, and formulate the control problem of minimizing state estimation error through the generation of optimal measurement bases. In Sec. IV, we present the estimation results for general mixed two- and three-level density matrices, comparing asymptotic to finite sample performance and interrogating the robustness of optimal estimation to control field noise. In the concluding Sec. V, we draw conclusions regarding the efficiency of system-theoretic methods for optimal quantum estimation.

II. ESTIMATION OF THE DENSITY MATRIX

A. Quantum estimation and the likelihood function for state reconstruction

Quantum statistical inference is based on the notion of a quantum probability space [1,24].

space dimension N [$O(2^n)$ for n -qubit systems], compared to the single measurement type required for such system theoretic strategies, which exploit the dynamical evolution of the system to extract state information.

¹The problem of optimal Hamiltonian parameter estimation [37] is distinct from that of optimal state estimation and is not addressed here.

²The number of measurement types required to reconstruct the state with standard state tomography scales as $O(N)$ with Hilbert

Definition 1. A quantum probability space is a pair $(\mathcal{N}, \rho(\theta))$, where \mathcal{N} is an operator algebra, typically the algebra of bounded operators $B(\mathcal{H})$ on a Hilbert space \mathcal{H} , and $\rho(\theta)$ is the positive, unit trace density operator (parameterized by a vector θ of parameters). For a finite-dimensional Hilbert space of dimension N , to which we restrict our attention, $\rho(\theta)$ is an $N \times N$ positive-semidefinite (hence Hermitian) matrix. The bounded operators we are concerned with are the Hermitian observables.

The probability distributions of measurement outcomes for any of a set of mutually commuting observables are determined by $\rho(\theta)$ together with a positive operator-valued measure. Consider a measurable space (χ, \mathcal{A}) , where χ is the set of possible outcomes (here, eigenvalues λ_i of any of the observable operators), and \mathcal{A} is the σ algebra of subsets of χ .

Definition 2. A positive operator-valued measure (POVM) is a map $M : \mathcal{A} \rightarrow B_+(\mathcal{H})$ [where $B_+(\mathcal{H})$ is the set of bounded positive operators on the Hilbert space \mathcal{H}] that is a resolution of the identity, i.e., for a finite-dimensional Hilbert space of dimension N , $\sum_i F(x_i) = I_N, F(x_i) \in B_+(\mathcal{H})$. It defines the probabilities of measurement outcomes $x_i \in \chi$ according to probability density function $p(x_i|\theta) = \text{Tr}[\rho(\theta)F(x_i)]$.

We consider projective measurements where the $F(x_i) = F_i$ are orthogonal projectors [43]. Then any quantum observable can be written

$$O = \sum_{i=1}^N \lambda_i F_i, \quad F_i = |i\rangle\langle i|,$$

where λ_i denotes an eigenvalue of O and $|i\rangle$ denotes the corresponding eigenvector. The measurement outcomes for observable O —the eigenvalues λ_i —are indexed by the integers $(1, \dots, N)$, and each is associated with an F_i through the POVM map M .

For projective measurements, the POVM elements $F_i = |i\rangle\langle i|$ constitute a basis for any observable that commutes with O . The associated observables are then said to belong to the same “measurement basis.” It is convenient to represent each measurement basis in terms of an $N \times N$ unitary matrix of common eigenvectors V of $\{F_1, \dots, F_N\}$. Since each POVM is a resolution of the identity, the subset of $N - 1$ projectors $\{F_1, \dots, F_{N-1}\}$, together with $\rho(\theta)$, fully characterizes the probability measure for observations in this basis.

A collection of POVMs is informationally complete if specification of their associated probability density functions (pdfs) $p(x_i|\theta)$ uniquely determines the density operator $\rho(\theta)$. Since the Hermitian matrix ρ is of unit trace, it is specified by $N^2 - 1$ parameters, and we write $\{F_i\}, i = 1, \dots, N^2 - 1$ for such an informationally complete set of POVMs. In order to estimate the density matrix by measurements of a collection of Hermitian observables, it must be possible to distinguish all of the POVM elements of an informationally complete set based on the measurement outcomes (eigenvalues). Given that the density matrix is a function of $N^2 - 1$ independent parameters—and each POVM is characterized by $N - 1$ elements—a collection of $N + 1$ POVMs, with associated measurement bases $V^{(r)}, 0 \leq r \leq N$, is the minimum number required for informational completeness [35]. In the current work, the data x consist of m_j measurement outcomes in each of $N + 1$ measurement bases with $\sum_{j=1}^{N+1} m_j = m$. The

measurement bases are constructed via optimal control, as discussed in Secs. III and IV.

Asymptotically efficient estimation [42] of quantum states can be achieved by incorporating all information in the probability density function for measurements into the estimation procedure. Let $x = (x_1, \dots, x_m)$ be an independent and identically distributed (*i.i.d.*) sample of size m from a population with probability density function $p(x|\theta)$, which depends on the unknown parameter vector θ whose true value is θ_0 . The joint density of the sample defined as a function of the unknown parameter vector θ , $L(\theta|x)$, is called the *likelihood function*.

We collect all the distinct parameters of the density matrix ρ into the $(N^2 - 1)$ -dimensional vector θ . The most convenient parametrization of $\rho(\theta)$ differs based on the state estimation method; various parametrizations are discussed in Sec. II D. The likelihood function for quantum state estimation is then

$$L(\theta|X) = \prod_{k=1}^m \text{Tr}[\rho(\theta)F_{i_k}], \quad (1)$$

where F_{i_k} denotes the POVM element obtained in the k th draw ($i = 1, \dots, N^2 - 1; k = 1, \dots, m$), and X denotes the set of all measurement outcomes. $L(\theta|X)$ may be interpreted as the probability of obtaining the set of observed outcomes for a given density matrix $\rho(\theta)$. The maximum likelihood estimator of the density matrix seeks to identify the admissible parameter vector θ at which this likelihood is maximal.

B. Quantum maximum likelihood estimation

Without an asymptotically efficient estimator [42], a system-theoretic state estimation scheme cannot be optimal. The maximum likelihood estimator is asymptotically efficient.

The value of the parameter vector that maximizes the likelihood function is called the ML estimator of θ :

$$\begin{aligned} \hat{\theta}_{\text{ML}}^m &= \arg \max_{\theta \in \Theta} L(\theta|x) \\ &= \arg \max_{\theta \in \Theta} \left(\prod_{i=1}^m p(x_i|\theta) \cdots p(x_m|\theta) \right), \end{aligned}$$

where Θ denotes the admissible parameter space. Typically, the logarithm of the likelihood function, $\ln L(\theta|x)$, is easier to maximize numerically because of its separability. By maximizing the log likelihood, the ML estimator minimizes the Kullback-Leibler distance between the estimated and true probability distributions.³

ML has several properties that make it an attractive frequentist estimation procedure:

(1) *Consistency*: An estimator $\hat{\theta}^m$ is *consistent* for the parameter θ (written as $\text{plim } \hat{\theta}^m = \theta_0$) if for every $\epsilon > 0$,

$$\lim_{m \rightarrow \infty} P_{\theta} \{ |\hat{\theta}^m - \theta_0| \geq \epsilon \} = 0.$$

The ML estimator is consistent: $\text{plim } \hat{\theta}_{\text{ML}}^m = \theta_0$.

³The Kullback-Leibler distance is given by $E_x \{ \ln [p(x|\theta_0)/p(x|\hat{\theta})] \}$, where $p(x|\theta_0)$ is the true distribution and $p(x|\hat{\theta})$ is the estimated distribution [44].

(2) *Invariance*: The ML estimator of $c(\theta)$ is $c(\hat{\theta}_{\text{ML}}^m)$, for a continuous and continuously differentiable function $c(\cdot)$.

(3) *Asymptotic Normality*: For a sequence of estimators $\hat{\theta}^m$, if $k_m(\hat{\theta}^m - \theta_0) \xrightarrow{d} N(0, \Sigma)$ as $m \rightarrow \infty$, where \xrightarrow{d} denotes convergence in distribution and k_m is any function of m , $\hat{\theta}^m$ is said to be k_m consistent for θ and has an asymptotic normal distribution with asymptotic covariance matrix Σ . The ML estimator is asymptotically normally distributed:

$$\sqrt{m}[\hat{\theta}_{\text{ML}}^m - \theta_0] \rightarrow \mathcal{N}[0, mI^{-1}(\theta_0)],$$

where $I(\theta_0) = -\text{E} \left[\frac{\partial^2 \ln L(\theta_0|x)}{\partial \theta \partial \theta'} \right]$.

$I(\theta_0)$ is called the expected Fisher information matrix. Note that the asymptotic covariance matrix of the ML estimator is a function of the unknown parameters. Alternative approaches exist for the consistent estimation of the expected Fisher information matrix, thereby providing feasible versions of the observed Fisher information matrix. The most commonly used estimator replaces the expected second derivatives matrix of the log likelihood function with its sample mean evaluated at the maximum likelihood estimates,

$$\hat{I}_1(\hat{\theta}_{\text{ML}}^m) = - \left[\frac{\partial^2 \ln L(\hat{\theta}_{\text{ML}}^m|x)}{\partial \theta \partial \theta'} \right]. \quad (2)$$

(4) *Asymptotically efficient*: A sequence of consistent estimators $\hat{\theta}^m$ is asymptotically efficient if $\sqrt{m}[\hat{\theta}^m - \theta_0] \xrightarrow{d} \mathcal{N}[0, mI^{-1}(\theta_0)]$, where $I(\theta) = -\text{E} \left[\frac{\partial^2 \ln L(\theta|x)}{\partial \theta \partial \theta'} \right]$; $I^{-1}(\theta_0)$ is called the Cramér-Rao lower bound (CRB) for consistent estimators.

Property 4 is the subject of the following classic lemma of frequentist inference.

Lemma 1. The eigenvalues of the covariance matrix of parameter estimates of a consistent frequentist estimator are bounded from below by the eigenvalues of $I^{-1}(\theta_0) = \{-\text{E} \left[\frac{\partial^2 \ln L(\theta_0|x)}{\partial \theta \partial \theta'} \right]\}^{-1}$. The maximum likelihood estimator asymptotically (i.e., in the limit of an infinite number of measurements) achieves this lower bound.

Because of properties 1–4 and the fact that hypothesis testing procedures based on ML estimators are uniformly most powerful (UMP),⁴ the ML estimation methodology is considered the most desirable among frequentist estimation techniques.

In quantum ML estimation, we aim to identify the maximum of the likelihood function (1) over the set of *admissible* density matrices, which are elements of the Bloch vector space B_{N^2-1} (see the Appendix). All parametrizations of the density matrix [such as the Bloch vector parametrization (see the Appendix)] require the imposition of constraints on the parameter vector θ ; these constraints are necessary for expression (1) to be a well-defined likelihood. Assuming the constraints on the parameter vector θ are of the general form

$a_j(\theta) \geq 0, j = 1, \dots, N$, the problem can be formulated in terms of the Lagrangian function

$$\mathcal{L}(\theta, \lambda, \gamma | x) = \ln \left[\prod_{k=1}^m \text{Tr} [\rho(\theta) F_{ik}] \right] + \sum_{j=1}^N \zeta_j [a_j(\theta) - \gamma_j^2], \quad (3)$$

where the first term is $\ln L(\theta|x)$ in the absence of constraints [i.e., $\rho(\theta)$ need not be an admissible density matrix and L need not be a well-defined likelihood], the γ_j denote slack variables ($\gamma_j = 0$ in the case of an equality constraint), and the ζ_j denote Lagrange multipliers. It is convenient to order the N constraints such that the first constraint enforces the unit trace of ρ , and the following $N - 1$ constraints enforce its positive semidefiniteness. Note that $L(\theta | x)$ is a well-defined likelihood function only in the presence of these constraints. For parametrizations where the unit trace constraint is implicit in the parametrization (such as the Bloch vector parametrization [45]), $\zeta_1 = 0$. We denote the vector of parameters as $(\theta, \zeta, \gamma) \equiv \mathbf{t}$. Finding the optimum corresponding to this Lagrangian entails searching for parameter vector \mathbf{t} that renders the gradient vectors $\nabla L(\theta)$ and a linear combination of $\nabla(a_j(\theta) - \gamma_j)$, $j = 1, \dots, N$ parallel. There are two common approaches to solving this problem: (1) minimization of the “sum of squares” (of the first-order conditions) function $\sum_i (\frac{\partial \mathcal{L}}{\partial t_i})^2$; (2) finding the roots of the system of nonlinear equations $\frac{\partial \mathcal{L}}{\partial \mathbf{t}} = 0$ using the Newton-Raphson (NR) method. In fact, methods (1) and (2) may be combined to produce a globally convergent NR algorithm.

The dimension of the parameter space for quantum state estimation increases quadratically with the Hilbert space dimension, necessitating the use of efficient parametrizations of the density matrix. Here, we employ the so-called Bloch vector parametrization (see the Appendix), where the probability of an observable outcome according to the Born rule, $\text{Tr}[\rho(\theta) F_i]$, is a simple linear function of the parameter vector θ . Moreover, the Bloch vector parametrization is perhaps the most commonly used in the statistical physics of finite-dimensional quantum systems (especially in quantum information applications). In the Bloch vector parametrization, $\theta_j = \text{Tr}(\lambda_j \rho), i = 1, \dots, N^2 - 1$, where the λ_j denote the generators of the Lie group $\text{SU}(N)$ (see the Appendix).

Note that the ML estimator obtained by maximizing the likelihood defined by the Lagrangian (3) is consistent, asymptotically normally distributed, and has an asymptotic covariance matrix equal to the inverse of the expected Fisher information matrix, $I^{-1}(\theta_0)$. We estimate the expected Fisher information using Eq. (2).

C. Asymptotically optimal measurements

The use of w ML guarantees asymptotic efficiency for a fixed set of measurements. However, since the likelihood function for measurements changes with time due to the evolution of the system, additional conditions must be satisfied by the measurement bases in order to minimize state estimation errors. Our aim is to achieve these conditions by appropriate choice of control inputs, rather than through modification of the measurement apparatus itself.

⁴The power of a test is the probability of rejecting a null hypothesis (such as $\theta = \theta_0$) given that it is false [42]; it is ideally close to 1. Such hypothesis tests based on ML estimators can be shown to be the most powerful in their respective classes.

In quantum statistics, there are multiple Cramér-Rao type inequalities, each with its own associated (quantum) Fisher information. Work in quantum probability theory [24] has indicated that $I^{-1}(\theta_0)$ for an arbitrary choice of measurement bases is generally not the tightest asymptotic lower bound achievable in quantum ML estimation. The measurements that maximize the Fisher information depend on the true, unknown state of the quantum system, which in present context would require control strategies that are conditional on the measurement data. Although the choice of measurement bases that can achieve the tightest possible Cramér-Rao bound depends on the true ρ , there exists an approach to optimal measurement that is agnostic to the true value of ρ . Wootters [35] proposed a construction of measurement bases that maximizes the *average* information (over the set of all possible density matrices) obtained via a set of m measurements. These so-called mutually unbiased measurement bases (MUB) are “maximally noncommutative” in the sense that a measurement in one basis provides no information as to the outcome of a measurement over a basis unbiased with respect to the current one. Let $I(\hat{\theta}, \rho(\theta_0))$ denote the observed Fisher information given the true state $\rho(\theta_0)$. MUB aims to maximize the average Fisher information over all possible $\rho(\theta_0)$ ’s:

$$I(\hat{\theta}) = \frac{1}{V_0} \int_{B_{N^2-1}} I(\hat{\theta}, \rho(\theta_0)) d\rho(\theta_0), \quad (4)$$

where $\Theta \subset B_{N^2-1}$ again denotes the admissible parameter space and V_0 is the volume of Θ , by an appropriate choice of measurement bases.⁵ It can be shown that this is equivalent to maximizing the average *Kullback-Leibler (KL) information gain* [44] upon updating the flat prior distribution to the asymptotic multivariate normal distribution. Maximizing the information gain is equivalent to minimizing the “uncertainty volume” in the parameter space; in the absence of measurements, this is equal to the volume of the Bloch vector space. The uncertainty distance for estimation of a single parameter is the standard deviation of the estimator; the uncertainty volume is the product of the standard deviations of the estimators for each of the parameters.

The total uncertainty volume is minimized when the subspaces of B_{N^2-1} associated with each of the measurement bases are mutually orthogonal. Wootters showed [35] that this condition is equivalent to requiring that

$$|\langle \mathbf{v}_i^{(r)}, \mathbf{v}_j^{(r')} \rangle| = \frac{1}{\sqrt{N}}, \quad (5)$$

where $\mathbf{v}_i^{(r)}, \mathbf{v}_j^{(r')}$ are column vectors in the bases $V^{(r)}, V^{(r')}$ respectively, and $|\langle \cdot, \cdot \rangle|$ denotes the modulus of the Hermitian inner product. Since the condition depends only on the modulus of the inner product, it is insensitive to the phases associated with the components of each eigenvector, a property that will be exploited below. Whereas mutual nonorthogonality may decrease the asymptotic uncertainty volume in particular subspaces of B_{N^2-1} , the total asymptotic uncertainty volume is always increased by such nonorthogonality. Explicit formulas

for measurement bases that satisfy (5) are known in the cases where the Hilbert space dimension N is the power of a prime, and are discussed and employed in Sec. III.⁶ Since MUB measurements maximize the average Fisher information, any control strategy that can drive the measurement bases for a given set of observables to be MUB will be optimal from the perspective of the Fisher information performance measure, as discussed below. If the estimator is asymptotically efficient, the overall state estimation procedure will be asymptotically optimal. These concepts are made rigorous below. In this paper we assess both asymptotic and finite sample optimality.

III. OPTIMAL OBSERVABILITY THEORY

A. Bilinear observability on B_{N^2-1}

Consider a quantum control system with field-free Hamiltonian H_0 and control Hamiltonian μ (e.g., molecular dipole operator), subject to single shaped electric or magnetic field input $\epsilon(t), t \in [0, T]$; generalization to multiple control inputs is straightforward. The goal is to optimally estimate the state at time $t = 0$ by proper choice of inputs and measurements over $[0, T]$. The time evolution of the state is governed by the von Neumann equation

$$\frac{d}{dt} \rho(t) = -\frac{i}{\hbar} [H_0 - \epsilon(t)\mu, \rho(t)], \quad \rho(0) = \rho_0. \quad (6)$$

As a control system with input function $\epsilon(\cdot)$, this system is said to be bilinear. Let the observables that can be measured be denoted $\{O_i\}$. Note these observables need not constitute a complete tomographic set. The time evolution of each of these observable operators is governed by the corresponding Heisenberg equations, with solution $O_i(t) = U^\dagger(t) O_i U(t)$, where

$$\frac{d}{dt} U(t) = -\frac{i}{\hbar} [H_0 - \epsilon(t)\mu] U(t), \quad U(0) = I. \quad (7)$$

The theory of observability provides necessary and sufficient conditions for the existence of a *consistent* estimator for all components of the state of a quantum dynamical system. For an observable system, any two states ρ^1, ρ^2 can be distinguished by expectation values of observations in the set $\{O_i\}$ along with controlled evolutions, i.e. $\text{Tr}[\rho^1(t) O_i] \neq \text{Tr}[\rho^2(t) O_i]$ for at least one O_i and t , where $\rho(t) = \rho_{\epsilon(\cdot)}(t, \theta)$. Here, we will be concerned with estimation of the initial state $\rho_0(\theta) \equiv \rho(0, \theta)$ of system (6), which we will denote by $\rho(\theta)$ for concision. Let $O'_i = O_i - \frac{\text{Tr} O_i}{N} I_N$, a traceless Hermitian operator. The observability condition [28] is that the rank of the direct sum of the subspaces S_i of $\text{su}(N)$ spanned by commutators of the form $[\dots [i H_{j_3}, [i H_{j_2}, [i H_{j_1}, i O'_i]]] \dots]$, where each $H_j \in \{H_0, \mu\}$, is equal to $N^2 - 1$, i.e.,

$$\text{rank} \bigoplus_{i=1}^n \text{span}\{[\dots [i H_{j_3}, [i H_{j_2}, [i H_{j_1}, i O'_i]]] \dots]\} = N^2 - 1,$$

⁵The Bures volume measure on the space of $N \times N$ density matrices is implied in Eq. (4).

⁶Another approach to optimal quantum measurement that does not depend on the true state of the system is symmetric informationally complete positive operator-valued measures (SIC-POVMs) [32]. These do not maximize the average Fisher information as do MUBs.

where n denotes the number of observable operators. Equivalently, system (6) is observable if and only if $\bigoplus_{i=1}^n \text{span}\{\dots [i H_{j_3}, [i H_{j_2}, [i H_{j_1}, i O'_i]] \dots]\} = \text{su}(N)$.

Identifiability of the state, given a set of measurements, can also be assessed using the theory of state tomography [16], which is a form of *method of moments* (MM) estimation. The tomographic inversion method estimates the parameters by equating population moments - here, the expectation values of the observable operators O_i - to the corresponding sample moments: the parameter estimates $\hat{\theta}_j, 1 \leq j \leq N^2 - 1$, are obtained by inverting a system of equations of the form

$$\text{Tr}[\rho(\hat{\theta})O_i] = c_i, \quad 1 \leq i \leq N^2 - 1, \quad (8)$$

where c_i denotes the sample mean of the observable quantity corresponding to measurement of O_i . Introducing the notation $A_{ij} = \frac{\partial \text{Tr}[\rho(\theta)O_i]}{\partial \theta_j}$, we may solve for the estimated parameter vector as $\hat{\theta} = A^{-1}\mathbf{c}$ for any parametrization $\rho(\theta)$ that is linear in θ (for example, the Bloch vector parametrization; see the Appendix). State identifiability corresponds to invertibility of the matrix A .

Now assume there are $n < N^2 - 1$ observables O_i , and write $O_i(t) = U^\dagger(t)O_i U(t), i = 1, \dots, n$ (for a continuous set of observation times $t \in [0, T]$), in the Heisenberg picture. We replace the data equation (8) by the n -component vector $y(t)$ of sample moments (properly corrected for the trace of the operators $\rho, \{O_i\}$) as a function of measurement time:

$$\text{Tr}[\rho'(\hat{\theta})U^\dagger(t)O'_i U(t)] = y_i(t), \quad 1 \leq i \leq n.$$

Let $A(t) = [\nu(O_1(t)), \dots, \nu(O_n(t))]$ denote an $(N^2 - 1) \times n$ -matrix function of time, where $\nu(B)$ for any Hermitian matrix B denotes the vector consisting of the $N^2 - 1$ independent parameters of the traceless Hermitian matrix $B - \frac{\text{Tr}(B)}{N}I_N$. The observation equation can then be written concisely as $A^T(t)\nu(\rho(\hat{\theta})) = y(t)$. Left multiplying by $A(t)$ and integrating over time, we have

$$\begin{aligned} \int_0^T A(t)A^T(t)dt\nu(\rho(\hat{\theta})) &= \int_0^T A(t)y(t)dt \\ \hat{\theta} &= \left[\int_0^T A(t)A^T(t)dt \right]^{-1} \int_0^T A(t)y(t)dt, \end{aligned} \quad (9)$$

where the second line follows if the parametrization is linear in θ and there are no constraints on the parameter vector. Then, identifiability is equivalent to the nonsingularity of the *observability Gramian* matrix $\int_0^T A(t)A^T(t)dt$ under the given evolution. Equation (9) is the continuous-time analog of the recursive least squares estimator [46] for θ :

$$\hat{\theta} = \nu(\hat{\rho}) = \left[\sum_{k=1}^p A(t_k)A^T(t_k) \right]^{-1} \left[\sum_{k=1}^p A(t_k)y(t_k) \right]. \quad (10)$$

Note that this approach requires a parametrization $\rho(\theta)$ that is linear in θ , in order for the Born observation law to be linear in the state parameters. The observability Gramian as written in (9) is a linear map $M : \text{su}(N) \rightarrow \text{su}(N)$, and hence the corresponding solution for $\hat{\theta}$ may violate the positive-semidefiniteness constraints on $\rho(\theta)$. If the system is observable, there exists at least one controlled trajectory such that the observability Gramian is nonsingular.

Observability theory is based on expectation values of observables, rather than the pdfs of the observations. As such, observability conditions can assess whether the system is identifiable (i.e., whether a consistent estimator exists) under some controlled evolution, but cannot say anything about the efficiency of a measurement scheme. Both tomography and the observability Gramian approaches minimize the least squares error between the model-predicted expectation values of the observations and the corresponding sample means. Since the observation law is in fact stochastic, the expectation value of each observation can at best be estimated in finite samples. Although dynamic methods require fewer observables to effectively create a complete tomographic set, all inversion approaches only use the information in a chosen set of moments of the data. Hence, unlike the ML estimator, such MM estimators are not asymptotically efficient for observation distributions that are not Gaussian (which they are not in the present case); their asymptotic variance does not attain the Cramér-Rao lower bound in Lemma 1 [47]. Additional conditions based on the Fisher information must be satisfied for optimality of systems theoretic state estimation schemes.

B. Optimal quantum observability

Whereas observability theory is concerned with identifiability of states of time-evolving quantum systems, the majority of work on quantum estimation *efficiency* has been restricted to measurements made at a single time. We introduce the notion of *optimal observability*, which is concerned with whether controlled observations $\{O_i(t)\}$ exist such that the Fisher information per observation of the measurement data, averaged over all possible true states ρ_0 , can be maximized.

Definition 3. A quantum system $\{H_0, \mu\}$ with observable operators $\{O_i\}$ is said to be *optimally observable* if there exists a sequence of controls $\{\epsilon_k(\cdot)\}$ and associated measurement times t_k such that $\max_{\{O_i, \epsilon_k(t_k)\}} \langle \|I(\rho_0(\hat{\theta}))\| \rangle / m = \max_{F_i} \langle \|I(\rho_0(\hat{\theta}))\| \rangle / m$, where m denotes the number of measurements, $\|\cdot\|$ denotes the Hilbert-Schmidt matrix norm, F_i are the elements of any informationally complete POVM set, and the averaging is in the sense of Eq. (4).

Let U_k denote the unitary propagator of the quantum system at a specified measurement time, evolving under the influence of a control field $\epsilon_k(\cdot)$ according to the Schrödinger equation. The measurement time may change with k , but without loss of generality we will assume it remains fixed at the value T , and that each control field is applied on the time interval $[0, T]$.

While observability theory generally assumes that only the expectation values of the operators O_i can be estimated, the MM formulation upon which it is based can be extended to accommodate the ability to distinguish among distinct measurement outcomes. Let us diagonalize the operators O_i as $Y_i \tilde{O}_i Y_i^\dagger$, where

$$\tilde{O}_i = \text{diag}(\underbrace{\lambda_1, \dots, \lambda_1}_{l_1}, \dots, \underbrace{\lambda_{q_i}, \dots, \lambda_{q_i}}_{l_{q_i}}),$$

and let $O_{ij} = Y_i \text{diag}(0, \dots, 0, \lambda_j, \dots, \lambda_j, 0, \dots, 0) Y_i^\dagger$. Outcomes in different degenerate subblocks j can be distinguished by measuring O_i . A necessary and sufficient condition for optimal observability is then that the following discrete-time

observability Gramian

$$\sum_{k=1}^{N+1} A_i^k (A_i^k)^T, \quad (11)$$

where $A_i^k = [v(U_k^\dagger O_{i1} U_k), \dots, v(U_k^\dagger O_{iq_i} U_k)]$, is nonsingular for some control sequence $\{\epsilon_k(\cdot)\}$ producing measurement bases $\{Y_{i_k}^\dagger U_k\}$ that satisfy the MUB conditions (5). A_i^k represents the measurement of observable O_i under controlled evolution k . Note that in order to satisfy the rank condition, it is necessary to measure a nondegenerate observable operator O_i (i.e., with $q_i = N$) for each controlled evolution. If multiple observables are required for nonsingularity, the system is said to display *weakly optimal observability*. This result accommodates the non-Gaussian distribution of measurement outcomes corresponding to the fact that $N + 1$ multinomials specify the probability distribution for quantum observations. Note, however, that the inverse Gramian should not be used to solve algebraically for parameter estimates, as this estimator is not asymptotically efficient. Moreover, in order to solve for ρ_0 algebraically while enforcing positive semidefiniteness, it is necessary to introduce a Riemannian metric [48] on the Bloch vector space, methods for which will be reported in a separate work.

A *time-independent sufficient* condition for optimal observability is that (i) the system is density matrix controllable [5], i.e., $\text{rank } \mathcal{L}\{H_0, \mu\} = N^2 - 1$, where \mathcal{L} denotes the dynamical Lie algebra generated by H_0 and μ (i.e., the span of their nested commutators), and (ii) $\{O_i\} = O_{\text{ND}}$, any fully nondegenerate observable. Condition (i) guarantees that the necessary measurement bases $\{Y_{i_k}^\dagger U_k\}$ can be generated by control. Controllability is not a necessary condition; it is not possible to specify a time-independent necessary condition based on linear algebraic rank, since the space of optimal measurement bases cannot be generated via a (linearly parameterized) vector space. Practically, the test for full density matrix controllability:

$$\text{span}\{[\dots [i H_{j_3}, [i H_{j_2}, i H_{j_1}]] \dots]\} = \text{su}(N),$$

can be applied to verify (i).

C. Optimal control theory for unitary measurement bases

The problem of steering the time evolution of the system to maximize the average Fisher information of the measurement data while minimizing control and measurement costs is one of multiobjective optimal control theory [49]. Increasing available measurement resources decreases the required control resources, and vice versa. In such problems the notion of optimality must be replaced by that of Pareto optimality: a control field $\epsilon^*(t)$ is said to be *Pareto optimal* if all other fields $\epsilon(t)$ have a greater value for at least one of the objective functions, or else have the same value for all objectives [39]. The set of Pareto optimal solutions is called the Pareto frontier. Here, the objectives are

$$J_1 = -\frac{1}{m} \langle \|I(\rho_0(\hat{\theta}))\| \rangle, \quad J_2 = \sum_{k=1}^N \frac{1}{2} \int_0^T \epsilon_k^2(t) dt.$$

This is a high dimensional optimization problem, possibly replete with local traps, which is greatly simplified by the

optimal observability theory. Replacing $-\langle \|I(\rho_0(\hat{\theta}))\| \rangle$ with the distances between the controlled measurement bases $\tilde{U}_k = \{Y_{i_k}^\dagger U_k\}$ and the optimal bases $V^{(k)}$, there are then N new objectives. The Hilbert-Schmidt (Frobenius) distances between the true $[\tilde{U}_k(T)]$ and target $[V^{(k)}]$ bases at time T :

$$\begin{aligned} F(\tilde{U}_k(T)) &= \|V^{(k)} - \tilde{U}_k(T)\|^2 \\ &= \text{Tr}\{[V^{(k)} - \tilde{U}_k(T)]^\dagger [V^{(k)} - \tilde{U}_k(T)]\} \\ &= 2N - 2\text{ReTr}\{[V^{(k)}]^\dagger \tilde{U}_k(T)\}, \end{aligned} \quad (12)$$

are used as the *Mayer costs* [46] that are to be minimized over the space of admissible controls, subject to the dynamical equation of motion (7). The unitary propagators corresponding to the MUB measurement bases jointly maximize the average Fisher information (FI); the first basis need not be generated since it is the identity at time $t = 0$. Prior work [40,50] has established regularity conditions under which search algorithms solving this control optimization problem will not encounter local traps. Control fields at which traps may exist are called singular extremals. Although it is always possible to find singular extremals [51], they are rarely encountered in the course of optimal control theory (OCT) calculations. Recent works have carefully assessed the frequency with which traps are encountered in unitary control landscapes [40,52]. These studies executed thousands of OCT runs with physically motivated Hamiltonians, including those of the type considered in the present work, and no traps were found.

The total fluence J_2 of the optimal control fields is the *Lagrange cost* $L(\tilde{\epsilon}(t))$ [4,40] that is to be minimized. The optimal control problem is then

$$\min_{\tilde{\epsilon}(\cdot)} \left\{ F(\tilde{U}_k(T)), \sum_{k=1}^N \frac{1}{2} \int_0^T \epsilon_k^2(t) dt \right\},$$

where \min is understood to denote a nondominated Pareto optimum. This is the simplest formulation of the problem, in terms of N separately controlled evolutions $U_k(T)$. For a single controlled evolution, the problem is

$$\min_{\epsilon(\cdot)} \left\{ F(\tilde{U}(t_k)), \frac{1}{2} \int_0^T \epsilon(t) dt \right\}.$$

If the system is fully controllable, a solution with just one controlled evolution is guaranteed to exist if the measurement times are chosen to be separated by at least the minimal controllability time [53]. For fully controllable systems the kinematic Pareto frontier [i.e., that on $U(N)$ [39]] consists of all $(N + 1)$ -tuples of \tilde{U}_k that satisfy the MUB condition. If the system is observable but not controllable, it is generally not possible to analytically characterize the Pareto frontier, but the discrete-time Gramian (11) can be applied to facilitate numerical sampling of the frontier using one or both of the following methods:⁷

⁷In single-control formulations, if the system is uncontrollable, given a control input function $\epsilon(t)$ obtained from the MUB basis control problem, one can generally increase the Fisher information per measurement by searching numerically for N -tuples of measurement times t_1, \dots, t_k that minimize one or both of these objectives.

(i) minimization of the condition number [54] of the discrete-time Gramian (11); (ii) minimization of the biasedness function $\sum_{j>i} \sum_{r>r'} |\langle \mathbf{v}_i^{(r)}, \mathbf{v}_j^{(r')} \rangle|$, where the \mathbf{v} are the measurement basis vectors. While the biasedness objective function is a more accurate indicator of the total Fisher information than the Gramian condition number, its minimization is not sufficient since it does not guarantee identifiability, unlike the latter's.

The associated optimal control problems can be solved using so-called homotopy tracking algorithms [49,55].⁸ These algorithms follow a specified track F_s of objective function values, where s denotes algorithmic time, toward the global minimum of each objective (12). The following differential equation specifies the evolution of each control field ϵ_k in continuous algorithmic time:

$$\frac{\partial \epsilon(s,t)}{\partial s} = f(s,t) + \frac{a(s,t)}{\gamma(s)} \left(\frac{dF_s}{ds} - \int_0^T a(s,t') f(s,t) dt \right).$$

Here $a(s,t)$ denotes the functional derivative (gradient) of the Hilbert-Schmidt distance with respect to the control $\epsilon(t)$, $\gamma(s) = \int_0^T a^2(s,t) dt$ is the norm square of the gradient, and $f(s,t)$ is a “free” function that arises due to the fact that the control problem is underdetermined in the absence of a Lagrange cost [49]. Once $\epsilon_k(t)$ are found that maximize the Fisher information per measurement, all other control fields corresponding to maximal information can be reached by applying the homotopy tracking algorithm above with $\frac{dF_s}{ds} = 0$. Then, each free function specifies a path parameterized by s through the space of control fields producing maximum Fisher information per measurement. In particular, the choice $f(s,t) = -\frac{\partial L(\epsilon(t))}{\partial \epsilon_k} |_{\epsilon(s,t)}$ minimizes the resource cost for the k th control, subject to the constraint $F(\tilde{U}_k) = 0$, thus enabling convergence to the Pareto frontier. For uncontrollable systems, this constraint can be replaced by $F(\tilde{U}_k) = \chi$, for some $\chi > 0$.

Since any $U_k(T)$ within the coset $SU(N)/T^{N-1}$ —where T^{N-1} denotes the maximal torus subgroup of $SU(N)$ [57]—produces an identical probability distribution of observations, each measurement basis control problem requires fewer resources and search effort than does the control of quantum gates on either $U(N)$ or $SU(N)$.⁹ For systems displaying weakly optimal observability (i.e., the observables O_i have degenerate spectra), control on lower-dimensional coset manifolds is required; for example, if each O_i is a pure state projector, only control on $\mathbb{C}\mathbb{P}^{N-1}$ is necessary. More generally, for a system where the observable O_i is of the form

$$Y_i \text{diag}(\underbrace{\lambda_1, \dots, \lambda_1}_{l_1}, \dots, \underbrace{\lambda_{q_i}, \dots, \lambda_{q_i}}_{l_{q_i}}) Y_i^\dagger,$$

only control on $U(N)/U(l_1) \times \dots \times U(l_{q_i})$ is required. However, for systems that display only weakly optimal observabil-

ity, J_1 will be n -fold greater than its value for an optimally observable system, and J_2 will also be strictly greater due to the need for more controlled evolutions.

For $N = 2$ systems, the mutually unbiased measurements we use in our simulations would correspond to measurements of the x, y , and z components of the spin of a single qubit if the measurements were all made at time $t = 0$, i.e., the $V^{(r)}, 0 \leq r \leq N$, can be written

$$\begin{aligned} V^{(0)} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, & V^{(1)} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \\ V^{(2)} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}. \end{aligned} \quad (13)$$

In the present approach, if one measures a single nondegenerate observable operator O_1 that commutes with the field-free Hamiltonian, the system would be optimally observable if it is possible to generate the following operators in the Heisenberg picture through controlled time evolutions:

$$\begin{aligned} O_1 &= \begin{pmatrix} \lambda_1 & \\ 0 & \lambda_2 \end{pmatrix}, & O_{1,\epsilon_2(\cdot)}(t_1) &= \frac{1}{2} \begin{pmatrix} \lambda_1 + \lambda_2 & \lambda_1 - \lambda_2 \\ \lambda_1 - \lambda_2 & \lambda_1 + \lambda_2 \end{pmatrix}, \\ O_{1,\epsilon_3(\cdot)}(t_2) &= \frac{1}{2} \begin{pmatrix} \lambda_1 + \lambda_2 & i(\lambda_2 - \lambda_1) \\ i(\lambda_1 - \lambda_2) & \lambda_1 + \lambda_2 \end{pmatrix}, \end{aligned}$$

where we have used the notation for controlled Heisenberg picture operators in Definition 3 and assumed $\epsilon_1(\cdot)$ is the zero field with $t_1 = 0$. As a second example, consider the case of $n = 2$ (two-qubit) systems that are fully controllable. For any Hilbert space of dimension $N = 2^n$ —which encompasses all multiqubit systems—the MUB bases can be expressed in terms of the eigenvectors of the Pauli operators (tensor product of Pauli matrices $\sigma_{x,y,z}$). Then measurements of any pair of Pauli operators from the set $\{(\sigma_z^1, \sigma_z^2), (\sigma_x^1, \sigma_y^2), (\sigma_y^1, \sigma_x^2), (\sigma_y^1 \sigma_y^2, \sigma_z^1 \sigma_x^2), (\sigma_x^1 \sigma_x^2, \sigma_y^1 \sigma_z^2)\}$ [58], in the Heisenberg picture at appropriate times under controlled evolutions, would permit state reconstruction with minimum uncertainty. Such a system would display weakly optimal observability.

To apply the treatment to state estimation of molecular systems, the Hilbert space dimension N may be approximated as any power of 2 or an odd prime such that $N \geq N'$, where N' denotes the effective dynamical dimension for modeling the system away from the dissociative limit. For $N = 3$, or more generally when N is the power of an odd prime, a canonical construction of the bases $V^{(r)}$ is [35]

$$V_{pq}^{(r)} = \begin{cases} \delta_{pq}, & r = 0 \\ \frac{1}{\sqrt{N}} \exp\left[\frac{2\pi i}{N}(rp^2 + pq)\right], & 1 \leq r \leq N. \end{cases} \quad (14)$$

N' is determined by defining a fraction σ of the population of the most populated molecular eigenstate; eigenstates with an initial population smaller than this are considered unpopulated for the purposes of the density matrix estimation. Two problems where state estimation of molecules is useful are (a) simultaneous excitation of multiple eigenstate populations [49], and (b) laser cooling of molecular internal degrees of freedom [59]. In the former case, the Boltzmann distribution provides an approximate lower bound on N' , whereas in the latter case it provides an upper bound on N' . An example is

⁸Dynamic optimization (optimal control) problems typically require specialized algorithms since the dynamical constraint is a differential equation that must be satisfied for each feasible control; homotopy tracking algorithms are ideal for multiobjective control problems [56].

⁹Detailed studies of the search complexity and resource scaling of gradient-based algorithms for gate control, including properties of the matrices H_0, μ that facilitate the search for optimal controls, are reported in [40].

the control of rotational transitions in diatomic molecules. The field-free Hamiltonian H_0 is

$$H_0 = \frac{\hbar^2}{2I} \sum_{J=0}^{N-1} J(J+1)|J\rangle\langle J|, \quad (15)$$

where J denotes the total angular momentum quantum number and $I = mr^2$ is the molecular moment of inertia. We assume $\epsilon(t)$ is oriented along the z axis and the Stark effect is negligible; the $|J\rangle$ states are $(2J+1)$ -fold degenerate. The selection rules for the system are $\Delta J = \pm 1$, $\Delta M = 0$ (with transition dipole matrix elements $\langle JM|\mu|(J+1)M\rangle = \mu_{\text{perm}}\sqrt{\frac{(J+M+1)(J-M+1)}{(2J+1)(2J+3)}}$, where μ_{perm} is the permanent molecular dipole moment [60]). In this example, setting $\sigma = 10^{-3}$, $N' \approx 10$ for rotational states of the HCl molecule in the Boltzmann distribution at 300 K, whereas for CO, $N' \approx 30$.

For a more realistic rovibrational model for diatomics, the overall Hilbert space dimension of the system is $N = N_J N_v$, where N_J and N_v denote the numbers of rotational and vibrational states included in the model, respectively. The rotational and vibrational degrees of freedom are separable, and a Morse oscillator model can be used to represent the vibrational degrees of freedom. The dipole operator μ will contain more nonzero elements due to the fact that there are no formally forbidden vibrational transitions, but an upper bound on N'_v is imposed by the finite number of bound vibrational states.

In either case, if these systems are controllable on the N -dimensional Hilbert space, they are optimally observable using measurements of the energy alone. The corresponding controlled POVM for a single observable O_{ND} that commutes with H_0 is then

$$\begin{aligned} F_{r(N-1)+i} &= V^{(r)} \tilde{F}_i (V^{(r)})^\dagger, \\ \tilde{F}_i &= |i\rangle\langle i| = \text{diag}(0, \dots, 1, \dots, 0), \\ 1 &\leq i \leq N-1. \end{aligned}$$

IV. SIMULATION RESULTS

In this Section we simulate optimal state estimators for two- and three-level quantum dynamical systems. For the

Hamiltonian operators in the control system (6), we choose the following two sets, which are controllable on $U(N)$ and allow only nearest neighbor transitions between energy eigenstates (all matrix elements are in atomic units):

$$H_0 = k \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mu = \begin{pmatrix} 1 & 0.3736 \\ 0.3736 & 1 \end{pmatrix}, \quad (16)$$

$$H_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.0088 & 0 \\ 0 & 0 & 0.0263 \end{pmatrix}, \quad (17)$$

$$\mu = \begin{pmatrix} 1 & 0.3736 & 0 \\ 0.3736 & 1 & 0.3736 \\ 0 & 0.3736 & 1 \end{pmatrix}.$$

Although the optimal bases listed in Sec. III C belong to the unitary group $U(N)$, any bases within the corresponding unitary cosets $SU(N)/T^{N-1}$ are physically equivalent since they produce identical probability distributions for measurements [57]. In particular, spin systems controllable on $SU(N)$ are optimally observable. However, such systems typically require multiple control fields to generate $SU(N)$. In order to enable a study of the scaling of the estimation accuracy and efficiency with Hilbert space dimension, both systems were designed to be controllable on $U(N)$ rather than $SU(N)$ through the use of operators with a nonzero trace, and with only one control.

Parameters for the $N=2$ system were aligned to the typical field strengths and evolution times for NMR systems, by setting $k = B_z M$ in (16), with appropriate choice of static field strength B_z and magnetic moment M (Fig. 1). The optimal measurements are computed according to (13), which is applicable to spin systems of Hilbert space dimension 2^n . The $N=3$ system, for which optimal measurements are computed according to (14), was chosen to represent molecular systems of arbitrary Hilbert space dimension. If the observable measured is the field-free Hamiltonian operator H_0 , the observations can be made after the field is turned off, since only the populations and not the relative phases matter. Thus for both $N=2$ and $N=3$, the nondegenerate observable is $O_{\text{ND}} = H_0$.

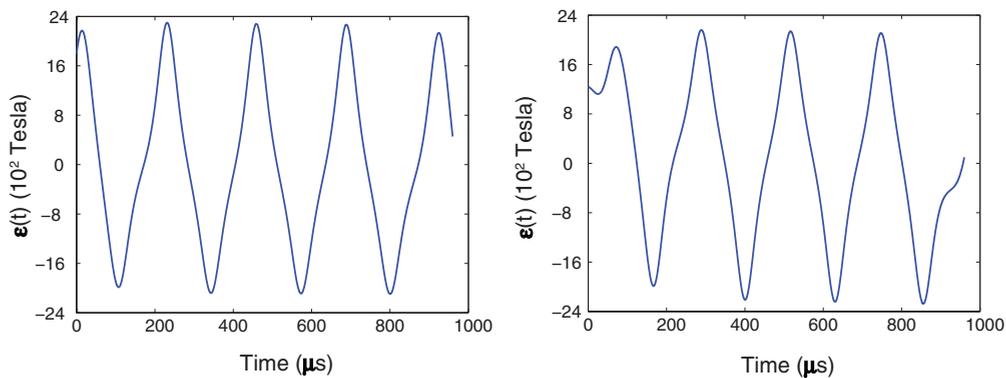


FIG. 1. (Color online) Optimal control fields $\epsilon(t)$ [see Eq. (6)] for driving the two-level system (16) to MUB measurement bases (13), for the choice $k = B_z M$, with B_z on the order of 10 T and M set to the proton magnetic moment. (Left) Field for basis $V^{(1)}$; (Right) Field for basis $V^{(2)}$. The Frobenius distance (12) between controlled and target bases is < 0.03 in both cases.

TABLE I. Asymptotic and finite sample distribution statistics (1000 repeated samples) for optimally controlled state estimation of two-level quantum systems: MUB measurement bases. The row labeled “asymptotic” in each panel reports the point estimates of the parameters and diagonal elements, along with the asymptotic standard errors in parentheses and asymptotic 95% confidence intervals in square brackets. (Since estimates for the off-diagonal coherences follow directly from those for θ , they are omitted for brevity.)

Panel A: Sample size 100					
	θ_1	θ_2	θ_3	ρ_{11}	ρ_{22}
True value	-0.44	-0.02	0.19	0.59	0.41
Asymptotic	-0.27 (0.15) [-0.56 -0.019]	-0.39 (0.12) [-0.64 -0.15]	0.39 (0.12) [0.15 0.64]	0.70 (0.06) [0.58 0.82]	0.30 (0.06) [0.18 0.42]
Bias	-0.40×10^2	-0.14×10^2	-0.55×10^2	-0.27×10^2	0.27×10^2
Standard error	0.16	0.19	0.17	0.09	0.09
RMSE	0.16	0.19	0.17	0.09	0.09
95% CE	[-0.76 -0.09]	[-0.39 0.33]	[-0.15 0.52]	[0.42 0.77]	[0.23 0.58]
Panel B: Sample size 1000					
	θ_1	θ_2	θ_3	ρ_{11}	ρ_{22}
Asymptotic	-0.35 (0.03) [-0.40 -0.30]	0.01 (0.05) [-0.10 -0.12]	0.20 (0.05) [0.10 0.29]	0.60 (0.023) [0.55 0.64]	0.40 (0.02) [0.36 0.45]
Bias	-0.46×10^2	0.00	0.36×10^2	0.18×10^2	-0.18×10^2
Standard error	0.06	0.07	0.07	0.04	0.04
RMSE	0.06	0.07	0.07	0.04	0.04
95% CE	[-0.56 -0.35]	[-0.14 0.09]	[0.09 0.32]	[0.54 0.77]	[0.22 0.46]

We consider several Monte Carlo simulation environments in order to assess the asymptotic and finite-sample properties of the ML estimators of the density matrix and test statistics for various physical quantities of interest. For both two- and three-level systems, we report results for full rank, nondegenerate mixed states. The true ρ 's (see Row 1 of Tables I and II) are randomly chosen and, hence, the results may be considered representative for any true underlying density matrix.¹⁰ For each ρ , 1000 hypothetical samples of i.i.d. quantum observations each of size¹¹ $m = 100, 400, 1000$ are simulated with the observations evenly distributed between the $N + 1$ controlled measurement bases. For simulating quantum observations from a given basis $V^{(r)}$, the multinomial distribution probabilities (p_1^r, \dots, p_N^r) are computed as $p_i^r = \text{Tr}[\rho(\theta)F_{ik}^r]$ where F_{ik}^r denotes the POVM element associated with the multinomial outcome i (hence eigenvalue λ_i of O_{ND}) obtained in draw k from basis r . The parameters of ρ are then estimated using the ML approach for each sample.

A. Optimally controlled MUB measurements

The primary goals in this Section are to (i) interrogate the efficiency of the optimal system-theoretic state reconstruction scheme; and (ii) determine whether the asymptotic normal

distribution of the parameter estimates provides a good approximation to the finite-sample distribution when optimal measurements are generated by controlled evolutions $\rho_{\epsilon(\cdot)}(t)$, and how the approximation improves with increase in the sample size.

1. Two-level systems

Figure 1 depicts the optimal control fields for driving the two-level system (16) to each of the MUB measurement bases (13). These fields were obtained via optimal control theory methods (Sec. III C) that produce shaped pulses. While bang-bang control solutions [43,61] that apply only one group generator at a time are well-known for some two-level propagators, such as $V^{(1)}$ (which is the Hadamard gate), for higher N such control strategies (also called uniform finite generation) must be solved numerically, and also require more control pulse energy and overall evolution time.

Table I reports statistics from the asymptotic and finite-sample distributions of the ML estimators of the parameters and diagonal elements of the density matrix for the mixed two-level system.¹² Panels A and B report results for $m = 100$ and 1000, respectively. In particular, we report the following statistics from the finite-sample distribution: bias, standard deviation (std), root mean square error (RMSE), and 95% confidence intervals.

Note that the diagonal elements of the density matrix are smooth real functions of the parameter vector θ . Let $\rho_{ii}(\theta)$ denote the i th diagonal element of the density matrix. Given the invariance property of the ML estimator (see Sec. II),

¹⁰The uniform probability measure on the θ parameter vector space was used to generate random density matrices, by acceptance-rejection sampling. We also considered choices of the true density matrix other than the ones reported in this section, and the results are not sensitive to the choice of the true ρ .

¹¹These choices of sample sizes enable us to assess the impact of the sample size on the properties of the parameter estimates for experimentally realistic scenarios.

¹²Since estimates for the off-diagonal coherences follow directly from those for θ , they are omitted for brevity.

TABLE II. Asymptotic and finite sample distribution statistics (1000 repeated samples) for optimally controlled state estimation of three-level quantum systems: MUB measurement bases. The row labeled “asymptotic” in each panel reports the point estimates of the parameters and diagonal elements, along with the asymptotic standard errors in parentheses and asymptotic 95% confidence intervals in square brackets.

Panel A: Sample size 100											
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	ρ_{11}	ρ_{22}	ρ_{33}
True value	0.15	-0.14	-0.07	-0.04	-0.15	-0.01	-0.17	-0.23	0.23	0.30	0.46
Asymptotic	0.12 (0.17) [-0.20 -0.45]	-0.37 (0.10) [-0.56 -0.18]	-0.16 (0.12) [-0.18 -0.14]	-0.09 (0.13) [-0.33 -0.16]	0.04 (0.15) [-0.26 0.35]	-0.04 (0.15) [-0.34 -0.25]	-0.26 (0.10) [-0.45 -0.07]	-0.41 (0.12) [-0.65 -0.17]	0.14 (0.05) [0.04 0.25]	0.28 (0.08) [0.13 0.44]	0.57 (0.07) [0.43 0.71]
Bias	-0.55×10^2	2.20×10^2	1.90×10^2	0.89×10^2	-0.22×10^2	-0.69×10^2	1.43×10^2	5.35×10^2	2.50×10^2	0.59×10^2	-3.09×10^2
Standard error	0.15	0.15	0.14	0.15	0.15	0.16	0.16	0.17	0.08	0.09	0.10
RMSE	0.15	0.15	0.14	0.15	0.15	0.16	0.16	0.17	0.09	0.09	0.10
95% CE	[-0.19 0.44]	[-0.39 0.18]	[-0.36 0.20]	[-0.31 0.29]	[-0.46 0.12]	[-0.33 0.28]	[-0.46 0.16]	[-0.47 0.16]	[-0.14 0.51]	[0.08 0.82]	[0.17 0.96]

Panel B: Sample size 1000											
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	ρ_{11}	ρ_{22}	ρ_{33}
Asymptotic	0.20 (0.04) [0.120 0.29]	-0.20 (0.04) [-0.28 -0.11]	-0.18 (0.001) [-0.19 -0.18]	-0.06 (0.05) [-0.15 -0.04]	-0.16 (0.04) [-0.24 -0.09]	0.05 (0.05) [-0.04 -0.15]	-0.15 (0.03) [-0.22 -0.08]	-0.13 (0.04) [-0.22 -0.05]	0.28 (0.03) [0.230 0.33]	0.30 (0.03) [0.250 0.37]	0.41 (0.03) [0.360 0.46]
Bias	0.04×10^2	-0.31×10^2	-0.01×10^2	-0.16×10^2	-0.32×10^2	0.07×10^2	-0.25×10^2	-0.21×10^2	-0.06×10^2	-0.06×10^2	0.12×10^2
Standard error	0.06	0.06	0.05	0.05	0.06	0.05	0.06	0.06	0.03	0.03	0.04
RMSE	0.06	0.06	0.05	0.05	0.06	0.05	0.06	0.06	0.03	0.03	0.04
95% CE	[0.04 0.27]	[-0.25 -0.03]	[-0.16 0.02]	[-0.15 0.06]	[-0.27 -0.05]	[-0.11 0.10]	[-0.29 -0.07]	[-0.36 -0.12]	[0.11 0.29]	[0.16 0.36]	[0.39 0.70]

the ML estimator of $\rho_{ii}(\theta)$ is $\rho_{ii}(\hat{\theta})$. Also, the asymptotic distribution of the estimators of the diagonal elements of ρ can be obtained using the continuous mapping theorem [42]: $\rho_{ii}(\hat{\theta})$ has an asymptotic normal distribution with asymptotic variance given by

$$\text{var}(\rho_{ii}(\hat{\theta})) = \left(\frac{\partial \rho_{ii}(\theta_0)}{\partial \theta_0} \right)^T \Sigma \left(\frac{\partial \rho_{ii}(\theta_0)}{\partial \theta_0} \right), \quad (18)$$

where Σ is the asymptotic covariance matrix of $\hat{\theta}$.

Row 1 of Panel A reports the true values of the parameters and the diagonal elements of the density matrix. The row labeled ‘‘Asymptotic’’ in each panel reports the point estimates of the parameters and diagonal elements, along with the asymptotic standard errors in parentheses and asymptotic 95% confidence intervals in square brackets. We estimate the asymptotic covariance matrix consistently using the observed Fisher information given by Eq. (2). As mentioned above, the asymptotic distribution of the estimators of the diagonal elements of ρ are obtained using the continuous mapping theorem. For the computation of the asymptotic distribution, 1 out of the 1000 hypothetical samples is selected randomly.

Due to the nonlinearity of the likelihood function (1), it is essential to interrogate the accuracy of the asymptotic ML predictions. For several of the parameters and diagonal elements, the asymptotic confidence intervals contain the true values of the parameters. The asymptotic standard errors decrease with the increase in sample size from $m = 100$ to $m = 1000$ in Panels A–C. Consequently, as predicted by asymptotic theory, the distributions get narrower with increase in the sample size.

The subsequent rows of each panel report statistics from the finite-sample distributions of the parameter estimates. The Table reveals that the finite-sample biases are negligible, even for small sample sizes, such as $m = 100$ in Panel A. The finite-sample standard errors closely track the asymptotic standard errors for most of the parameters. Consequently, the finite-sample confidence intervals display similar coverage to the asymptotic ones. Inference based on the asymptotic distribution of the parameters can be reliable for as small as 100 total measurements (25 per controlled evolution).

Panels (a)–(c) in Fig. 2 plot asymptotic and finite-sample distributions of $\sqrt{m}[\rho_{11}(\hat{\theta}) - \rho_{11}(\theta_0)]$ (the upper diagonal element of the density matrix), for $m = 100, 400$, and 1000 , respectively. Here, the asymptotic distribution is nondegenerate. Note again that the finite sample distributions closely mirror the asymptotic distributions for all three sample sizes. The quality of the normal approximation to the finite sample distribution increases with sample size.

2. Three-level systems

The analysis so far has been restricted to two-level quantum systems. Most proposed applications of state estimation, including coupled spin states and molecular electronic, rotational, and vibrational states, involve higher dimensional density matrices. In order to assess the impact of the Hilbert space dimension on the performance of optimally controlled state estimation, we investigate the state estimation of three-level ($N = 3$) systems. Figure 3 depicts the optimal control

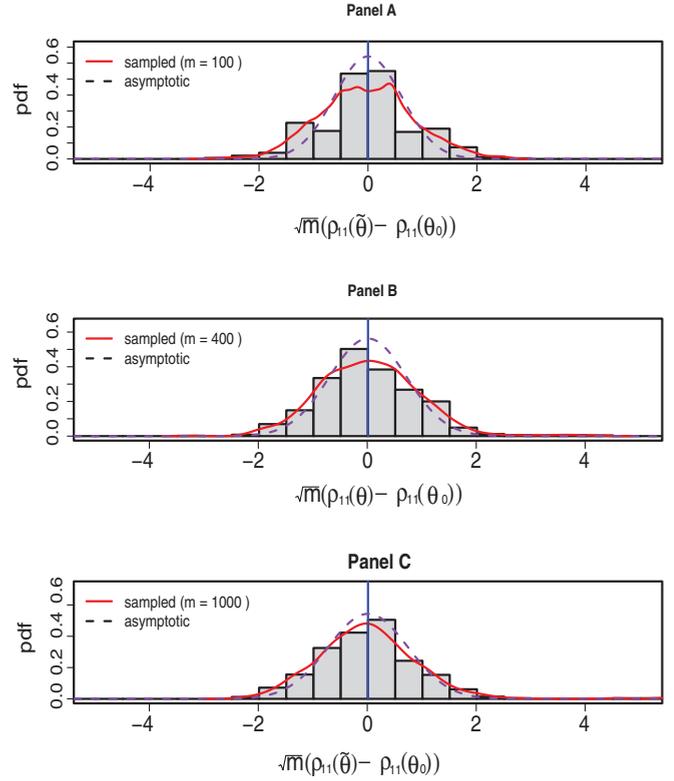


FIG. 2. (Color online) Finite sample distributions of $\sqrt{m}[\rho_{11}(\hat{\theta}) - \rho_{11}(\theta_0)]$, for mixed two-level state, MUB bases. (a) $m = 100$; (b) $m = 400$; (c) $m = 1000$. In each panel, finite sample distributions (1000 simulations) are shown alongside the corresponding asymptotic distribution.

fields, along with their power spectra, for driving the three-level system (17) to each of the MUB measurement bases (14).

Table II reports the asymptotic and finite-sample behavior of the ML estimators of the parameters and diagonal elements of the density matrix for a three-level quantum system. As in Table I, Panels A and B report results for $m = 100$ and 1000 , respectively. Note that the divergence between the asymptotic and finite-sample performance of the estimators is slightly more pronounced for some parameters in the three-level system, potentially because of the increase in the number of parameters and nonlinearity of the model. Note that the asymptotic standard errors decrease somewhat faster than the finite sample errors, providing an indication of how finite sample simulation is required to determine the rate of improvement of estimator accuracy with sample size. Still, the finite sample standard errors are close to their asymptotic counterparts. The small sample standard errors decrease with increase in the sample size—the standard errors in Panel B are about 1/3 of those in Panel A. This is also shown in Fig. 4, where each panel plots the finite sample distribution of the ML estimator of one parameter for sample sizes $m = 100, 400$, and 1000 in the same graph.

In addition to the asymptotic and finite-sample standard errors decreasing with sample size and, consequently, the confidence intervals becoming shorter, the quantity \sqrt{m} times the standard error is of similar magnitude across sample sizes—indicating that these sample sizes are close to the

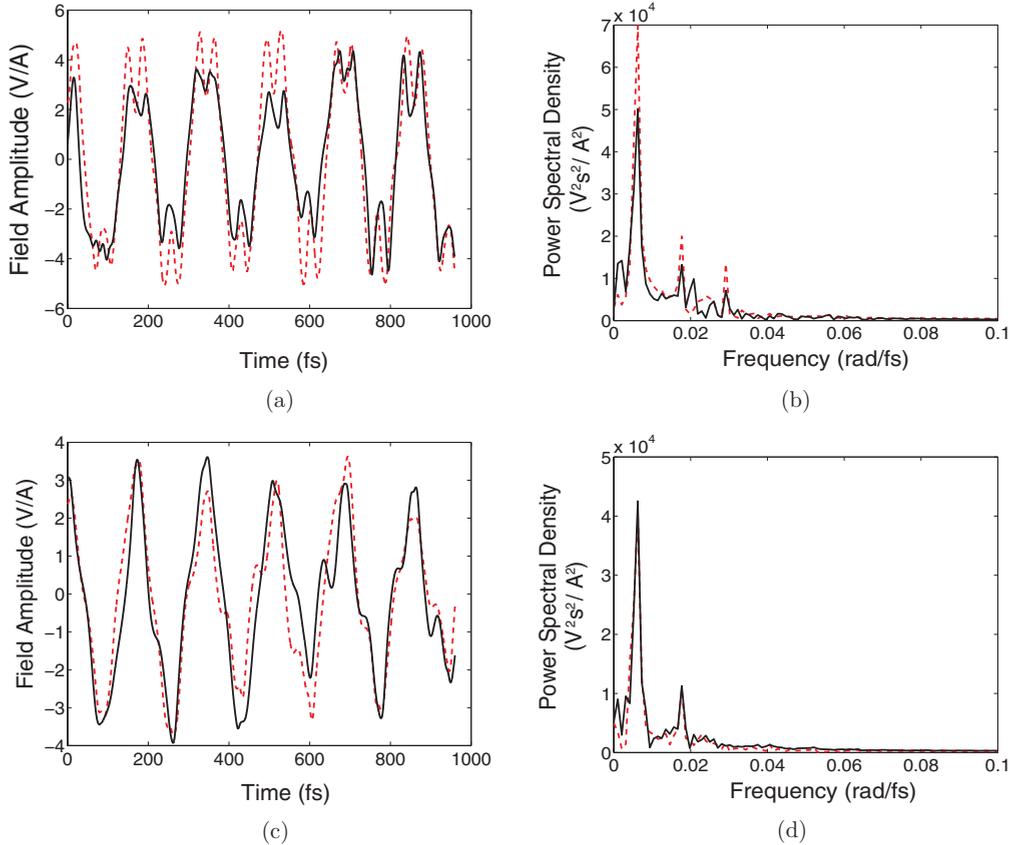


FIG. 3. (Color online) Optimal fields $\varepsilon(t)$ [see Eq. (6)] and power spectra (black, solid) for driving the three-level system (17) to MUB measurement bases (14). Noisy fields and spectra are superimposed (red, dashed). (a) Field for basis $V^{(1)}$; (b) power spectrum for basis $V^{(1)}$; (c) field for basis $V^{(2)}$; and (d) power spectrum for basis $V^{(2)}$. The Frobenius distance (12) between controlled and target bases is <0.03 for both optimal fields in the absence of noise.

asymptotic regime wherein the predictions of frequentist estimation theory—including the rate of convergence to the asymptotic limit—are most accurate. This is revealed in Fig. 5, which plots the finite-sample and asymptotic distributions of $\sqrt{m}[\rho_{ii}(\hat{\theta}) - \rho_{ii}(\theta_0)]$. Panels (a)–(c) plot the distributions for ρ_{11} for $m = 100, 400$, and 1000 , respectively, while panels (d)–(f) report the same for ρ_{22} . Again, the quality of the normal approximation to the parameter estimate distributions improves considerably with sample size.

B. Robustness of optimal state estimation to control field noise

In order for all the parameters of the density matrix to be identifiable by frequentist inference, measurements in at least $N + 1$ bases are required. In system-theoretic reconstruction strategies, proper choice of measurement times is essential to ensure identifiability. Completely random bases corresponding to $N + 1$ random measurement times during a single evolution were found to render the state unidentifiable, since they did not produce a full rank discrete time observability Gramian in Eq. (9) [which is equivalent to the set of POVM elements $F_{r(N-1)+i}, 1 \leq i \leq N - 1, 1 \leq r \leq N + 1$ spanning the Lie algebra $\mathfrak{su}(N)$]. However, the observability Gramian's rank cannot be used to distinguish between optimal and suboptimal measurement times or control strategies.

The results in the previous subsection were all obtained for simulated samples employing noise-free controls that precisely produced the MUB measurement bases maximizing the average Fisher information, or, equivalently, minimizing the asymptotic covariance matrix of the estimators, over the set of all possible density matrices. To undertake an analysis of the effect of suboptimal control on estimator efficiency, we simulated noise in the fields. It is crucial to undertake a sensitivity analysis to determine whether controlled measurement bases generated using noisy fields achieve similar estimation accuracies in finite samples, and to interrogate the advantages of optimal estimation. The primary goal of this subsection is therefore to compare the asymptotic and finite sample relative efficiencies of estimation strategies employing perfect and noisy controls. We assume here that the dynamical parameters are well known and that control errors are completely measurable, since only classical measurements of the field amplitude are required for the latter; analogous simulations could be carried out for the Hamiltonian parameter uncertainty.

The use of suboptimal noisy control induces the MUB bases to rotate away from their ideal unbiased configurations, causing the average Fisher information to decrease. After such an effective rotation, the new bases can be written

$$\tilde{V}^{(r)} = U(s)V^{(r)}, \quad (19)$$

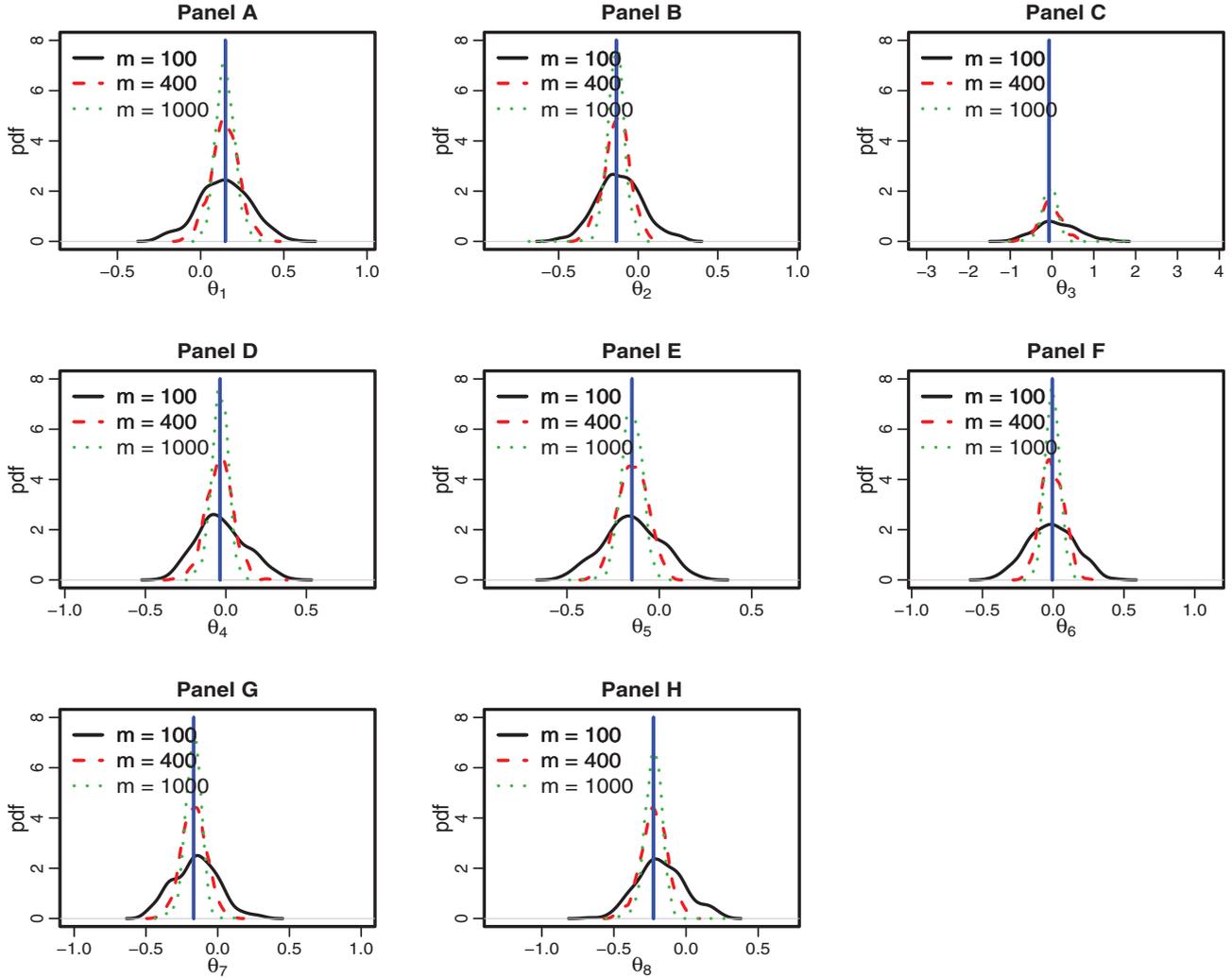


FIG. 4. (Color online) Finite sample distributions of parameter estimates, three-level system, MUB bases. Each panel reports results for one parameter of the density matrix and superimposes results from estimations using sample sizes 100, 400, and 1000. The finite sample distributions were computed from 1000 simulations.

where $U(s) = e^{iA^{(r)}s}$, A being a Hermitian matrix specifying a random axis of rotation in the N -dimensional Hilbert space, induced by the noise; s is a scalar parameter specifying the extent of rotation (magnitude of the solid angle). It can be shown [54] that for control system (7), $iAs = -\frac{i}{\hbar}V \int_0^T \mu(t)\delta\epsilon(t)dtV^\dagger$, where $\mu(t)$ is the time-evolved control Hamiltonian in the Heisenberg picture and $\delta\epsilon(t)$ is the noisy control increment. In the present case, for basis r , the amplitude of the field noise was chosen such that the inner product in (5) satisfied

$$|\langle \mathbf{v}_i^{(r)}, \mathbf{v}_j^{(r')} \rangle| \geq \alpha \frac{1}{\sqrt{N}}, \quad (20)$$

where r' runs over all the other bases, and α is a chosen scalar that is greater than unity ($\alpha = 1.2$ in the present work, for at least one pair of eigenvectors i, j from bases r and r' , respectively). (For generality in these simulations, optimal basis $V^{(0)} = I$ was also assumed to be generated by controlled evolution at time T , and the associated control field was

subjected to noise as well.) These systems were identifiable, as verified by the observability Gramian rank condition in Sec. III, but did not produce a valid optimal observability Gramian (11). The noisy fields are superimposed on the optimal fields in Fig. 3.

Table III reports the asymptotic and finite-sample performance of the ML estimators using the suboptimal bases for $N = 3$. Panels A and B report results for $m = 100$, and 1000, respectively. We begin with a comparison of the bias of the estimators for the optimal and suboptimal measurements. Comparing the results in Table III to those in Table II for optimal measurements, we find that the bias of the ML estimators for the suboptimal measurements are between one and two orders of magnitude larger than the corresponding values for the optimal measurements, for all sample sizes. Next we compare the asymptotic relative efficiencies ($RE_{1,2} = \frac{\sigma_2^2}{\sigma_1^2}$) of perfect and noisy control. Comparing with the results obtained with perfect control in Table II reveals that the asymptotic standard errors for the suboptimal measurement

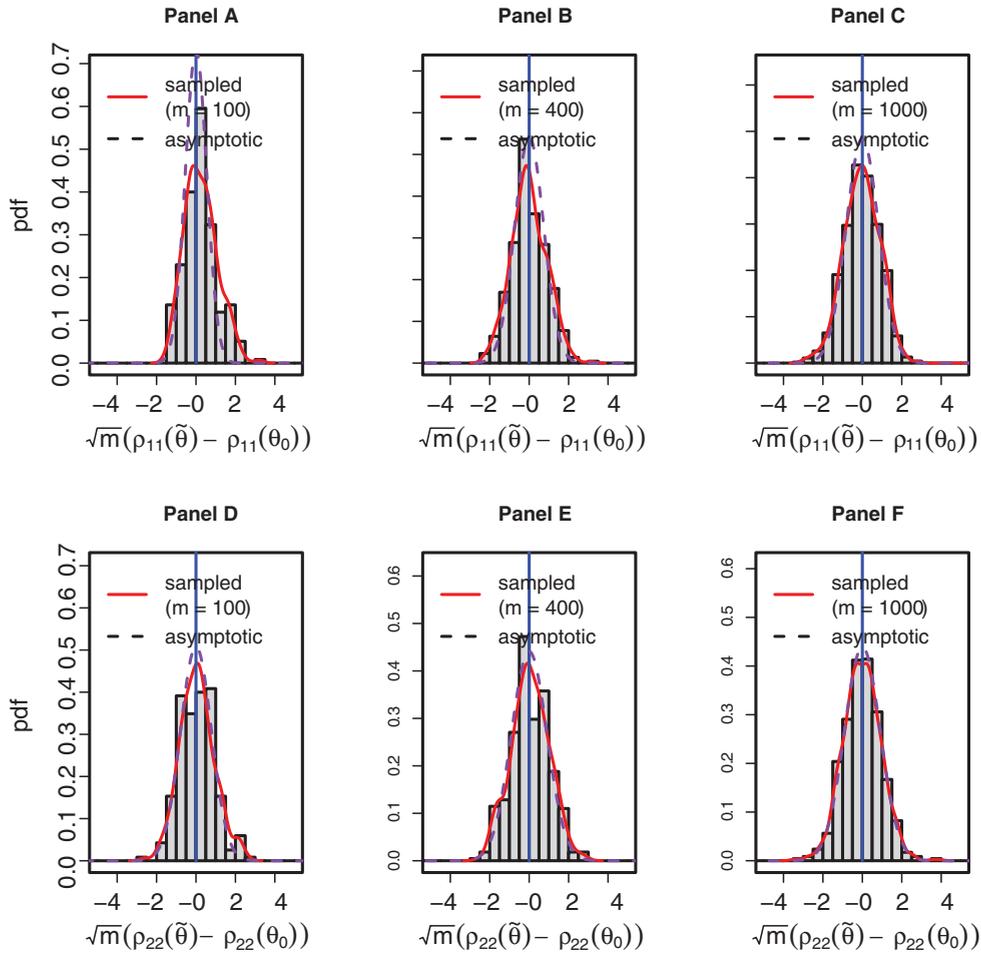


FIG. 5. (Color online) Finite sample distributions of $\sqrt{m}[\rho_{ii}(\hat{\theta}) - \rho_{ii}(\theta_0)]$, for mixed three-level state, MUB bases. Panels (a)–(c), ρ_{11} . (a) $m = 100$; (b) $m = 400$; and (c) $m = 1000$. Panels (d)–(f), ρ_{22} . (d) $m = 100$; (e) $m = 400$; and (f) $m = 1000$. In each panel, finite sample distributions (1000 simulations) are shown alongside the corresponding asymptotic distribution.

strategy are (considerably) bigger than those for the optimal strategy for most parameters, as expected based on the asymptotic theory. We next turn to a comparison of the relative finite-sample efficiencies of estimators using these measurement strategies. Table III reveals that the finite-sample standard errors for the parameters and diagonal elements of the density matrix are generally bigger for suboptimal controls. The divergence between finite sample standard errors of the suboptimal and optimal strategies increases for larger sample sizes. Thus, although the optimal state reconstruction schemes using perfect controls also have superior finite-sample properties than those using noisy controls, the difference can be marginal for small sample sizes and typical control field noise amplitudes. While the rate of convergence to the asymptotic limit is smaller than the theoretically predicted \sqrt{m} in both cases, it is greater for the perfect controls. Figure 6 compares asymptotic and finite-sample properties of the optimal MUB and the suboptimal bases for $m = 1000$. The normal approximation is less accurate for suboptimal measurements (but improves with increasing sample size).

To summarize, the results in this Section demonstrate that optimal measurements display superior asymptotic and finite sample properties for all sample sizes. The advantage increases for larger sample sizes, consistent with the asymptotic result

that optimal measurements achieve the maximal average Fisher information. We have interrogated the impact of control field noise, which can arise in laboratory applications, on the performance of optimal measurement strategies. The results indicate that the efficiency of optimally controlled state estimation is sufficiently robust to control field noise for small sample sizes.

V. CONCLUSION AND OUTLOOK

In this paper, we have shown how to make optimal quantum inferences based on a combination of control and measurement. We have developed a theoretical framework, called optimal observability theory, for assessment of the ability to reconstruct a quantum state with maximum accuracy despite restrictions on the measurement types available to the experimentalist, given a specification of available control resources. Methods have been presented for achieving these bounds in the case that they are achievable. Control and measurement resource requirements for optimal estimation have been quantified and minimized through the use of these methods. We have shown, through examples motivated by both NMR systems and model molecular systems, that with this technique it is possible to achieve reconstruction errors

TABLE III. Asymptotic and finite sample distribution statistics (1000 repeated samples) for optimally controlled state estimation of three-level quantum systems: noisy control fields. The row labeled “asymptotic” in each panel reports the point estimates of the parameters and diagonal elements, along with the asymptotic standard errors in parentheses and asymptotic 95% confidence intervals in square brackets.

Panel A: Sample size 100											
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	ρ_{11}	ρ_{22}	ρ_{33}
True value	0.15	-0.14	-0.07	-0.04	-0.15	-0.01	-0.17	-0.23	0.23	0.30	0.46
Asymptotic	0.11 (0.30) [-0.48 0.70]	0.28 (0.30) [-0.31 0.87]	0.15 (0.20) [-0.24 0.54]	-0.10 (0.20) [-0.49 0.29]	0.21 (0.20) [-0.18 0.60]	0.15 (0.40) [-0.63 0.93]	0.21 (0.30) [-0.80 0.38]	-0.21 (0.30) [-0.80 0.38]	0.27 (0.20) [-0.12 0.66]	0.13 (0.10) [-0.07 0.33]	0.60 (0.10) [0.40 0.80]
Bias	-13.5×10^2	23.9×10^2	6.41×10^2	3.31×10^2	11.6×10^2	23.9×10^2	4.98×10^2	8.27×10^2	5.59×10^2	-0.82×10^2	-4.77×10^2
Standard error	0.19	0.24	0.13	0.16	0.22	0.26	0.20	0.19	0.10	0.07	0.11
RMSE	0.23	0.34	0.14	0.16	0.25	0.35	0.20	0.20	0.11	0.07	0.12
95% CE	[-0.37 0.35]	[-0.42 0.49]	[-0.24 0.25]	[-0.39 0.27]	[-0.55 0.37]	[-0.25 0.66]	[-0.48 0.24]	[-0.52 0.21]	[-0.02 0.59]	[0.05 0.54]	[0.09 0.74]

Panel B: Sample size 1000											
	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	ρ_{11}	ρ_{22}	ρ_{33}
Asymptotic	0.09 (0.064) [-0.03 0.21]	-0.26 (0.032) [-0.32 -0.20]	-0.09 (0.032) [-0.152 0.032]	-0.10 (0.032) [-0.162 -0.038]	-0.13 (0.098) [-0.32 0.06]	0.08 (0.098) [-0.11 0.27]	-0.21 (0.098) [-0.40 -0.02]	-0.26 (0.098) [-0.45 -0.07]	0.21 (0.065) [0.08 0.34]	0.30 (0.033) [0.24 0.36]	0.48 (0.033) [0.42 0.54]
Bias	-7.87×10^2	14.7×10^2	1.48×10^2	-1.58×10^2	8.48×10^2	19.2×10^2	-1.32×10^2	2.56×10^2	1.48×10^2	-0.00	-1.48×10^2
Standard error	0.13	0.21	0.06	0.07	0.16	0.29	0.09	0.08	0.05	0.03	0.04
RMSE	0.15	0.26	0.06	0.07	0.18	0.35	0.09	0.08	0.05	0.03	0.05
95% CE	[-0.19 0.31]	[-0.35 0.39]	[-0.16 0.05]	[-0.19 0.10]	[-0.37 0.24]	[-0.33 0.63]	[-0.37 -0.01]	[-0.36 -0.06]	[0.11 0.34]	[0.26 0.36]	[0.36 0.58]

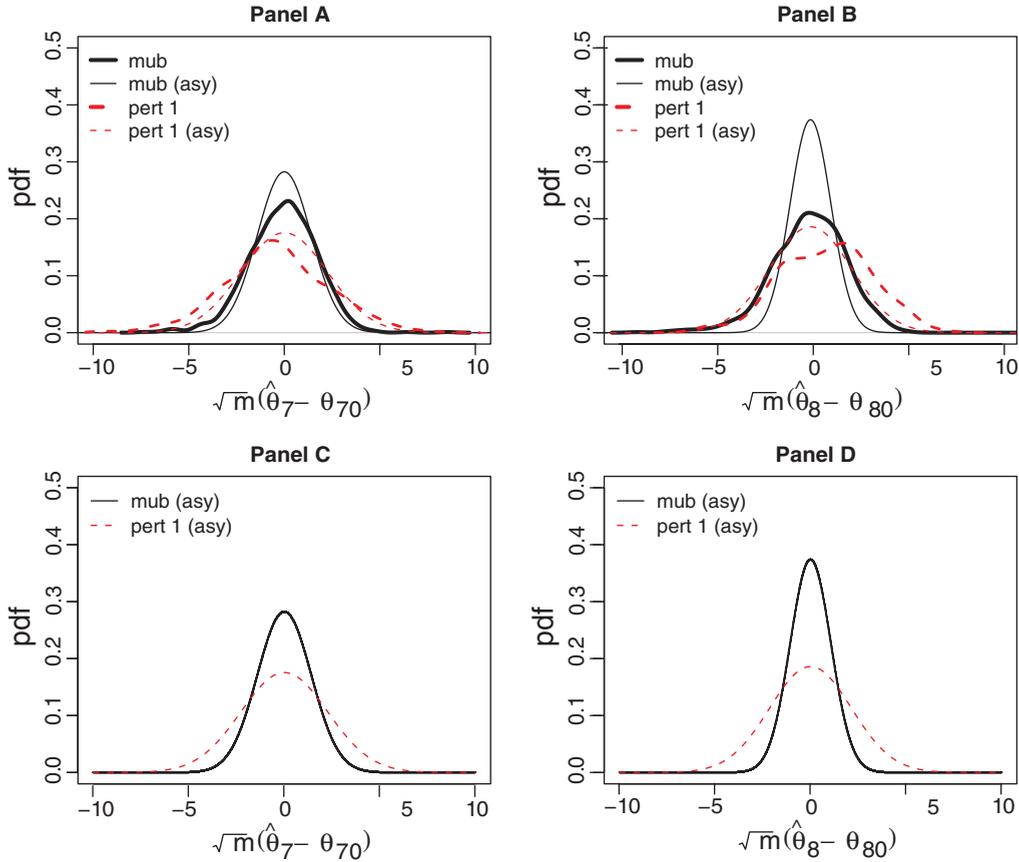


FIG. 6. (Color online) Distributions of $\sqrt{m}(\hat{\theta}_i - \theta_{i,0})$ for a three-level system, sample size 1000: comparison of perfect and noisy controls. (a) $\sqrt{m}(\hat{\theta}_7 - \theta_{7,0})$; (b) $\sqrt{m}(\hat{\theta}_8 - \theta_{8,0})$; and (c), (d) Asymptotic distributions obtained for (a) and (b), respectively, are isolated for clarity.

that are comparable to those of state-of-the-art quantum state reconstruction methods, but with significantly fewer measurement resources, even in the presence of noise in the electromagnetic fields used to manipulate system dynamics.

We have also demonstrated that these techniques provide more accurate state estimates than previously reported systems engineering schemes for state reconstruction. The finite sample performance of the optimal estimation strategy closely mirrors its asymptotic performance for sample sizes ≥ 25 measurements per controlled evolution time, and asymptotic errors are minimized due to the use of asymptotically efficient ML (rather than tomographic) methods. Optimal estimation efficiency is robust to field noise for smaller sample sizes. For larger sample sizes, further improvements in efficiency can be achieved by optimal estimation if control errors are reduced. Feedforward control techniques for doing so have been described [54].

Recent work [40] has considered the search complexity and resource scaling with respect to Hilbert space dimension for the optimal control of quantum unitary propagators. Those results can be used together with the present findings on the scaling of estimator efficiency in finite samples and the robustness of efficiency to control field noise, in order to assess the laboratory feasibility of the proposed estimation procedure for state reconstruction of particular atomic and molecular systems.

Regarding the outlook for the future development of optimal observability theory, first, even though the proposed scheme has been shown to be asymptotically optimal, and

the asymptotic standard errors closely mirror those in finite samples, the standard errors are often large. This feature is inherent to frequentist estimation schemes (since ML is the most efficient frequentist estimator). Further improvement of finite sample performance of quantum state estimation can be achieved through the use of Bayesian methods [10,62].

Second, while the MUB measurement bases applied in this work maximize the Fisher information of the measurement data averaged over all possible states ρ_0 , more efficient measurement strategies exist for any given state, due to the fact that the quantum Fisher information [1,3,23] of a measurement depends on the true state. In the context of stationary state estimation, the measurement resources required to implement such optimal strategies are highly restrictive due to the need to tailor the measurement strategy to every new state [3]. However, it is straightforward to generalize the system-theoretic optimal state reconstruction schemes introduced herein to such measurements. Since control resources are flexible and easily updated, one can adapt the applied control fields conditional on the outcomes of prior measurements, while keeping the measurements and the measurement times fixed. Moreover, the Pareto optimal control problem formulations presented in Sec. III C can be generalized to include systems wherein there are inequality constraints on the control (constrained optimal control [46,63]) or where the total evolution time rather than the control field fluence is to be minimized (time-optimal control [61]).

Finally, if the system Hamiltonian is unknown or imprecisely known, the present estimation strategy can be applied to simultaneously estimate those parameters as well as state parameters optimally, through an adaptive scheme [64] wherein (i) controls are computed based on the current Hamiltonian estimate; (ii) the resulting state estimates are used to update the Hamiltonian parameters, and (iii) the new Hamiltonian parameters are used to refine the choice of controls that maximize the extracted Fisher information. The resulting underdetermined (ill-posed) Hamiltonian estimation problem in (ii) can be rendered identifiable using Bayesian methods [10,62].

APPENDIX: BLOCH VECTOR PARAMETRIZATION

In the Bloch vector parametrization [45], the Hermitian operator ρ is parameterized in terms of an orthogonal basis $\{\lambda_j\}, 1 \leq j \leq N^2 - 1$, for the vector space of traceless Hermitian operators on an N -dimensional Hilbert space. In two dimensions, these are the familiar Pauli spin matrices, whereas in three dimensions they are the so-called Gell-Mann matrices. ρ can then be written

$$\rho \equiv \rho(\theta) = \frac{1}{N} I_N + \frac{1}{2} \sum_{j=1}^{N^2-1} \theta_j \lambda_j,$$

$$(\theta_1, \dots, \theta_{N^2-1}) \equiv \theta \in B_{N^2-1} \subset \mathbb{R}^{N^2-1},$$

where the $N^2 - 1$ matrices λ_j satisfy the conditions (a) $\lambda_j = \lambda_j^\dagger$, (b) $\text{Tr}(\lambda_j) = 0$, and (c) $\text{Tr}(\lambda_i \lambda_j) = 2\delta_{ij}$. These are the defining conditions of the generators of the Lie group $SU(N)$ that generalize the Pauli spin matrices. The θ_j are given by $\theta_j(\rho) = \text{Tr}(\lambda_j \rho)$ (i.e., are expectation values of the observable generators). The vector $\theta_j \lambda_j$ is called the Bloch vector.

B_{N^2-1} is a compact convex subset of \mathbb{R}^{N^2-1} . Let $a_i(\theta)$ denote the coefficients of the characteristic polynomial of ρ , $\det(yI_N - \rho)$, where ρ takes the form above. The unit trace constraint is automatically satisfied in the Bloch vector parametrization. It can be shown that the conditions of Hermiticity and positive semidefiniteness of ρ correspond to the following definition of the ‘‘Bloch vector set’’ B_{N^2-1} of admissible values of θ :

$$B_{N^2-1} \equiv \{\theta \in \mathbb{R}^{N^2-1} | a_i(\theta) \geq 0, i = 1, \dots, N\}.$$

This follows from the standard result that the roots of a characteristic equation are positive semidefinite if and only if the coefficients of the polynomial are positive semidefinite [45]. The a_i in the above definition of B are themselves polynomials in θ whose coefficients can be expressed in terms of the structure constants of the Lie algebra $\mathfrak{su}(N)$ of traceless skew-Hermitian matrices.

[1] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
 [2] J. D. Malley and J. Hornstein, *Stat. Sci.* **433**, 433 (2003).
 [3] O. E. Barndorff-Nielsen, R. D. Gill, and P. E. Jupp, *J. R. Statist. Soc.* **65**, 775 (2003).
 [4] C. Brif, R. Chakrabarti, and H. Rabitz, *New J. Phys.* **12**, 075008 (2010).
 [5] D. D’Alessandro, *Introduction to Quantum Control and Dynamics* (Taylor and Francis, Boca Raton, FL, 2007).
 [6] P. Brumer and M. Shapiro, *Annu. Rev. Phys. Chem.* **43**, 257 (1992).
 [7] S. A. Rice and M. Zhao, *Optical Control of Molecular Dynamics* (Wiley, New York, 2000).
 [8] U. Leonhardt, H. Paul, and G. M. D’Ariano, *Phys. Rev. A* **52**, 4899 (1995).
 [9] M. G. Raymer, M. Beck, and D. F. McAlister, *Phys. Rev. Lett.* **72**, 1137 (1994).
 [10] R. Derka, V. Buzek, G. Adam, and P. L. Knight, *J. Fine Mech. Opt.* **11–12**, 341 (1996).
 [11] G. Breitenbach, S. Schiller, and J. Mlynek, *Nature (London)* **487**, 471 (1997).
 [12] V. Buzek, *Ann. Phys.* **266**, 454 (1998).
 [13] A. M. Childs, I. L. Chuang, and D. W. Leung, *Phys. Rev. A* **64**, 012314 (2001).
 [14] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, *Phys. Rev. A* **64**, 052312 (2001).
 [15] A. G. White, A. Gilchrist, G. J. Pryde, J. L. O’Brien, M. J. Bremner, and N. K. Langford, *J. Opt. Soc. Am. B* **24**, 172 (2007).
 [16] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, *Phys. Rev. A* **77**, 032322 (2008).
 [17] K. Banaszek, G. M. D’Ariano, M. G. A. Paris, and M. F. Sacchi, *Phys. Rev. A* **61**, 010304 (1999).
 [18] Z. Hradil, *Phys. Rev. A* **55**, R1561 (1997).
 [19] M. Jezek, J. Fiurasek, and Z. Hradil, *Phys. Rev. A* **68**, 012305 (2003).
 [20] J. Rehacek, Z. Hradil, E. Knill, and A. I. Lvovsky, *Phys. Rev. A* **75**, 042108 (2007).
 [21] J. Rehacek, D. Mogilevtsev, and Z. Hradil, *New J. Phys.* **10**, 043022 (2008).
 [22] R. B. A. Adamson and A. M. Steinberg, *Phys. Rev. Lett.* **105**, 030406 (2010).
 [23] C. W. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
 [24] C. W. Holevo, *Statistical Structure of Quantum Theory* (Springer, Berlin, 2002).
 [25] P. B. Slater, *Nature (London)* **367**, 328 (1994).
 [26] K. R. W. Jones, *Ann. Phys. (NY)* **207**, 140 (1991).
 [27] A. Mansson, P. G. L. Mana, and G. Bjork, e-print arXiv:qwant-ph/0612105.
 [28] D. D’Alessandro, *J. Phys. A* **36**, 9721 (2003).
 [29] D. D’Alessandro and R. Romano, *Quantum Inf. Proc.* **5**, 139 (2006).
 [30] R. R. Mohler, *Bilinear Control Processes* (Academic Press, New York, 1973).
 [31] D. C. Brody and L. P. Hughston, *Phys. Rev. Lett.* **77**, 2851 (1996).
 [32] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, *J. Math. Phys.* **45**, 2171 (2004).
 [33] M. D. de Burgh, N. K. Langford, A. C. Doherty, and A. Gilchrist, *Phys. Rev. A* **78**, 052122 (2008).

- [34] O. E. Barndorff-Nielsen and R. D. Gill, *J. Phys. A* **33**, 4481 (2000).
- [35] W. K. Wootters and B. D. Fields, *Ann. Phys. (NY)* **191**, 363 (1989).
- [36] S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [37] Z.-M. Lu and H. Rabitz, *J. Phys. Chem.* **99**, 13731 (1995).
- [38] M. Demiralp and H. Rabitz, *Phys. Rev. A* **47**, 809 (1993).
- [39] R. Chakrabarti, R. Wu, and H. Rabitz, *Phys. Rev. A* **78**, 033414 (2008).
- [40] K. Moore, R. Chakrabarti, G. Riviello, and H. Rabitz, *Phys. Rev. A* **83**, 012326 (2011).
- [41] J. P. Palao and R. Kosloff, *Phys. Rev. A* **68**, 062308 (2003).
- [42] G. Casella and R. L. Berger, *Statistical Inference* (Duxbury Press, Pacific Grove, CA, 2002).
- [43] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
- [44] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (Chapman and Hall, New York, 2004).
- [45] G. Kimura, *Phys. Lett. A* **314**, 339 (2003).
- [46] R. Stengel, *Optimal Control and Estimation* (Dover, New York, 1994).
- [47] R. Engle and D. McFadden, *Handbook of Econometrics* (North Holland, Amsterdam, 1994).
- [48] M. S. Byrd and P. B. Slater, *Phys. Lett. A* **3**, 152 (2001).
- [49] R. Chakrabarti, R. Wu, and H. Rabitz, *Phys. Rev. A* **77**, 063425 (2008).
- [50] H. Rabitz, M. Hsieh, and C. Rosenthal, *Phys. Rev. A* **72**, 052337 (2005).
- [51] A. N. Pechen and D. J. Tannor, *Phys. Rev. Lett.* **106**, 120402 (2011).
- [52] K. W. Moore and H. Rabitz (unpublished).
- [53] V. Jurdjevic and H. J. Sussmann, *J. Diff. Equ.* **12**, 313 (1972).
- [54] R. Chakrabarti, *J. Phys. A* **44**, 185306 (2011).
- [55] J. Dominy and H. Rabitz, *J. Phys. A: Math. Theor.* **41**, 205305 (2008).
- [56] C. Hillermeier, *Nonlinear Multiobjective Optimization: A Generalized Homotopy Approach* (Birkhäuser, Basel, 2001).
- [57] T. Tilma and E. C. G. Sudarshan, *J. Phys. A* **35**, 10467 (2002).
- [58] J. Lawrence, C. Brukner, and A. Zeilinger, *Phys. Rev. A* **65**, 032320 (2002).
- [59] D. J. Tannor, R. Kosloff, and A. Bartana, *Faraday Discuss.* **113**, 365 (1999).
- [60] P. Atkins and R. Friedman, *Molecular Quantum Mechanics* (Oxford University Press, Oxford, 2005).
- [61] N. Khaneja, R. Brockett, and S. J. Glaser, *Phys. Rev. A* **63**, 032308 (2001).
- [62] P. B. Slater, *Physica A* **223**, 167 (1996).
- [63] R. Chakrabarti, H. Rabitz, S. L. Springs, and G. L. McLendon, *Phys. Rev. Lett.* **100**, 258103 (2008).
- [64] C. Brif and H. Rabitz, *J. Phys. B* **33**, L519 (2000).