IDENTIFICATION OF QUANTUM SYSTEMS: MAXIMUM LIKELIHOOD AND OPTIMAL EXPERIMENT DESIGN FOR STATE TOMOGRAPHY

Robert L. Kosut [*,1] **Ian A. Walmsley** [**,1] **Herschel Rabitz** [***,1]

[*] *SC Solutions, Sunnyvale, CA, USA,* kosut@scsolutions.com
[**] *Oxford University, Oxford, UK*
[***] *Princeton University, Princeton, NJ, USA*

Abstract:
An optimal experiment design problem invoked by the Cramér-Rao Inequality is applied to the problem of quantum state tomography. The optimization problem is integer-combinatorial and we use an established relaxation which results in a convex programming problem whose solution can be used to guide a more efficient experiment.

Keywords: system identification, experiment design, quantum systems

1. INTRODUCTION

The estimation of the quantum state of a system from available measurements is generally referred to as *quantum state tomography*. The more encompassing procedure of *system identification* has been referred to as *quantum process tomography*. There are many papers in the literature on the subject of quantum state and process tomography, *e.g.*, to name a few, (Nielsen and Chuang 2000), (James *et al.* 2001), (Grice and Walmsley 1996), (Paris *et al.* 2001), (Verstraete *et al.* 2001), (Walmsley and Waxer 1998). As shown originally in (Paris *et al.* 2001), *maximum likelihood estimation* (MLE) is directly applicable to quantum state tomography of a quantum system with non-continuing discrete measurements, *i.e.*, data is taken from repeated identical experiments at a finite number of (discrete) sample times. It is shown there that the MLE of the density matrix can be cast as a convex optimization problem. For continuous measurements a general formulation is presented in (Verstraete *et al.* 2001). In (Kosut, Rabitz, and Walmsley 2003) MLE is used as the identification step for adaptive control design.

[1] Supported by the DARPA QUIST Program

In this paper we address the related problem of experiment design so as to secure an estimate of the best quality. The approach presented relies on minimizing the Cramer-Rao lower bound where the design parameters are the number of experiments to be performed while the system is in a specified configuration. We solve the optimal experiment design problem using a standard relaxation method which results in a convex optimization problem. A more extensive treatment of this subject is in (Kosut, Walmsley, and Rabitz 2005). A completely data-based approach using learning control is presented in (Phan and Rabitz 1997); this is very valuable for quantum chemistry where models are not available.

2. QUANTUM STATE TOMOGRAPHY

Consider a quantum system which has $n_{\rm out}$ distinct *outcomes*, labeled by the index $\alpha,\ \alpha = 1, \ldots, n_{\rm out}$, and which can be externally manipulated into $n_{\rm cfg}$ distinct *configurations*, labeled by the index $\gamma,\ \gamma = 1, \ldots, n_{\rm cfg}$. Configurations can include wave-plate angles for photon counting, sample times at which measurements are made, and settings of any experimental "knobs" such as external control variables, *e.g.*, laser wave shape parameters, magnetic field strengths, and

so on. Although not discussed here, for quantum process tomography or Hamiltonian parameter estimation, configurations can also include a set of prepared initial states.

The problem addressed in this paper is to determine the minimum number of experiments per configuration in order to obtain a state estimate of a specified quality, *i.e.*, what is the tradeoff between number of experiments per configuration and estimation quality. The method used to solve this problem is based on minimizing the size of the Cramér-Rao lower bound on the estimation error (Cramér 1946).

2.1 *Data Collection*

The data is collected by recording measurement outcomes from identical experiments repeated ℓ_γ times in each configuration γ. Let $n_{\alpha\gamma}$ denote the number of times outcome α is obtained from the ℓ_γ experiments. Thus,

$$\sum_\alpha n_{\alpha\gamma} = \ell_\gamma, \quad \ell_{\rm expt} = \sum_\gamma \ell_\gamma \tag{1}$$

where $\ell_{\rm expt}$ is the total number of experiments. The *data set* consists of all the outcome counts,

$$D = \left\{ n_{\alpha\gamma} \;\middle|\; \begin{array}{l} \alpha = 1, \ldots, n_{\rm out} \\ \gamma = 1, \ldots, n_{\rm cfg} \end{array} \right\} \tag{2}$$

The design variables used to optimize the experiment are the non-negative integers $\{\ell_\gamma\}$ represented by the vector,

$$\ell = [\ell_1 \cdots \ell_{n_{\rm cfg}}]^T \tag{3}$$

Let $p^{\rm true}_{\alpha\gamma}$ denote the true probability of obtaining outcome α when the system is in configuration γ with unknown true state input $\rho^{\rm true}$. Thus,

$$\mathbf{E}\; n_{\alpha\gamma} = \ell_\gamma p^{\rm true}_{\alpha\gamma} \tag{4}$$

where the expectation $\mathbf{E}(\cdot)$ taken with respect to the underlying quantum probability distributions.

We pose the following *model* of the system, [2]

$$p_{\alpha\gamma}(\rho) = {\rm Tr}\; M_{\alpha\gamma}\sigma_\gamma(\rho) \tag{5}$$

where $p_{\alpha\gamma}(\rho)$ is the outcome probability of measuring α when the system is in configuration γ with input state ρ belonging to the set of density matrices,

$$\left\{ \rho \in \mathbf{C}^{n\times n} \mid \rho \geq 0,\; {\rm Tr}\;\rho = 1 \right\} \tag{6}$$

$\{M_{\alpha\gamma}\}$ are the POVM [3] elements of the measurement apparatus, and thus, for $\gamma = 1, \ldots, n_{\rm cfg}$,

$$\sum_\alpha M_{\alpha\gamma} = I_n, \quad M_{\alpha\gamma} \geq 0 \tag{7}$$

and $\sigma_\gamma(\rho)$ is the reduced density output state of the $\mathcal{Q}$-system model. A general (model) representation of a quantum system is the *Kraus operator-sum-representation* (OSR) (Nielsen and Chuang 2000). Specifically, in configuration γ, the system model can be parametrized by the set of *Kraus matrices*, $\{K_{\gamma k}\}$ as follows:

$$\sigma_\gamma(\rho) = \sum_{k=1}^{\kappa_\gamma} K_{\gamma k}\rho K_{\gamma k}^*, \; \sum_{k=1}^{\kappa_\gamma} K_{\gamma k}^* K_{\gamma k} = I_n \tag{8}$$

with $\kappa_\gamma \leq n^2$.

System in the model set We make the following assumption throughout: *the true system is in the model set*. This means that,

$$p^{\rm true}_{\alpha\gamma} = p_{\alpha\gamma}(\rho^{\rm true}) = {\rm Tr}\; O_{\alpha\gamma}\rho^{\rm true} \tag{9}$$

This is always a questionable assumption and in most engineering practice is never true. The case when the system is *not* in the model set will not be explored any further here except for the effect of measurement noise which is discussed next. It is important to emphasize that in order to produce an accurate unbiased estimate of the true density it is necessary to know the noise elements (as described next) which is a consequence of assumption (9).

Noisy measurements Sensor noise can engender more noisy outcomes than noise-free outcomes. Consider, for example, a photon detection device with two photon-counting detectors. If both are noise-free, meaning, perfect efficiency and no dark count probability, then, provided one photon is always present at the input of the device, there are only two possible outcomes: $\{10,\ 01\}$. If, however, each detector is noisy, then either or both detectors can misfire or fire even with a photon always present at the input. Thus in the noisy case there are *four* possible outcomes: $\{10,\ 01,\ 11,\ 00\}$.

Let $\{\, M_{\alpha\gamma} \mid \alpha = 1, \ldots, n_{\rm out} \,\}$ denote the noisy POVM and let $\{\, \overline{M}_{\alpha\gamma} \mid \alpha = 1, \ldots, \overline{n}_{\rm out} \,\}$ denote the noise-free POVM with $n_{\rm out} \geq \overline{n}_{\rm out}$ where,

$$M_{\alpha\gamma} = \sum_{\beta=1}^{\overline{n}_{\rm out}} \nu_{\alpha\beta\gamma}\, \overline{M}_{\beta\gamma} \tag{10}$$

The $\{\nu_{\alpha\beta\gamma}\}$ represents the noise in the measurement, specifically, the conditional probability that α is measured given the noise-free outcome β with the system in configuration γ. Since $\sum_\alpha\ \nu_{\alpha\beta\gamma} = 1,\ \forall\beta,\gamma$, it follows that if the noise-free set is a POVM then so is the noisy set.

---

[2] The notation here follows (Nielsen and Chuang 2000, §2) which provides an overview of the postulates of quantum mechanics.

[3] Positive Operator Valued Measure – a generalization of the quantum measurement process (Nielsen and Chuang 2000, §2.26).

2.2 *Maximum Likelihood State Estimation*

The Maximum Likelihood (ML) approach to quantum state estimation presented in this section, as well as observing that the estimation is convex, can be found in (Paris *et al.* 2001), (Verstraete *et al.* 2001) and the references therein. Using convex programming methods, such as an interior-point algorithm for computation, was not exploited in these references.

If the experiments are independent, then the probability of obtaining the data (2) is a product of the individual model probabilities (5). Consequently, for an *assumed* initial state ρ, the model predicts that the probability of obtaining the data set (2) is given by,

$$\mathbf{Prob}\,\{D,\rho\} = \prod_{\alpha,\gamma} p_{\alpha\gamma}(\rho)^{n_{\alpha\gamma}} \tag{11}$$

The data is thus captured in the outcome counts $\{n_{\alpha\gamma}\}$ whereas the model terms have a ρ-dependence. The *maximum likelihood estimate* (MLE) of ρ is obtained by finding a ρ in the set (6) which maximizes the *log-likelihood function*, or equivalently, minimizes the *negative log-likelihood function*,

$$L(D,\rho) = -\sum_{\alpha,\gamma} n_{\alpha\gamma} \log \mathrm{Tr}\, O_{\alpha\gamma}\rho \tag{12}$$

Specifically,

$$\begin{array}{l} \text{minimize } \; L(D,\rho) \\ \text{subject to } \; \rho \geq 0,\; \mathrm{Tr}\,\rho = 1 \end{array} \tag{13}$$

$L(D,\rho)$ is a positively weighted sum of log-convex functions of ρ, and hence, is a log-convex function of ρ. The constraint that ρ is a density matrix forms a convex set in ρ. Hence, (13) is in a category of a class of well studied log-convex optimization problems, *e.g.*, (Boyd and Vandenberghe 2004).

2.3 *Experiment Design for State Estimation*

In this section we describe the experiment design problem for quantum state estimation. We would like to select the number of experiments per configuration, the elements of the vector $\ell = [\ell_1 \cdots \ell_{n_{\mathrm{cfg}}}]^T \in \mathbf{R}^{n_{\mathrm{cfg}}}$, so as to minimize the error between the state estimate, $\widehat{\rho}(\ell)$, and the true state ρ^{true}. Specifically, we would like to solve for ℓ from:

$$\begin{array}{l} \text{minimize } \; \mathbf{E}\,\|\widehat{\rho}(\ell) - \rho^{\mathrm{true}}\|_{\mathrm{frob}}^2 \\ \text{subject to } \; \displaystyle\sum_{\gamma} \ell_\gamma = \ell_{\mathrm{expt}} \\ \qquad\qquad \text{integer } \ell_\gamma \geq 0 \end{array} \tag{14}$$

where ℓ_{expt} is the desired number of total experiments. This is a difficult, if not insoluble problem for several reasons. First, the solution depends on the estimation method which produces $\widehat{\rho}(\ell)$. Secondly, the problem is integer combinatoric because ℓ is a vector of integers, and most likely not convex. And finally, the solution depends on ρ^{true}, the very state to be estimated. Fortunately all these issues can be alleviated or circumvented.

We first eliminate the dependence on the estimation method. The following result can be established using the *Cramér-Rao Inequality* (Cramér 1946). [4]

State estimation variance lower bound

Suppose the system generating the data is in the model set used for estimation, i.e., *(9) holds. For $\ell = [\ell_1 \cdots \ell_{n_{\mathrm{cfg}}}]$ experiments per configuration, suppose $\widehat{\rho}(\ell)$ is a density matrix and an unbiased estimate of ρ^{true},* i.e., *$\widehat{\rho}(\ell) \geq 0$, $\mathrm{Tr}\,\widehat{\rho}(\ell) = 1$, and $\mathbf{E}\,\widehat{\rho}(\ell) = \rho^{\mathrm{true}}$. Under these conditions, the estimation error variance satisfies,*

$$\mathbf{E}\,\|\widehat{\rho}(\ell) - \rho^{\mathrm{true}}\|_{\mathrm{frob}}^2 \geq V(\ell,\rho^{\mathrm{true}}) \tag{15}$$

where

$$\begin{aligned} V(\ell,\rho^{\mathrm{true}}) &= \mathrm{Tr}\; G(\ell,\rho^{\mathrm{true}})^{-1} \\ G(\ell,\rho^{\mathrm{true}}) &= \sum_{\gamma} \ell_\gamma G_\gamma(\rho^{\mathrm{true}}) \in \mathbf{R}^{n^2-1\times n^2-1} \\ G_\gamma(\rho^{\mathrm{true}}) &= \sum_{\alpha} \frac{a_{\alpha\gamma} a_{\alpha\gamma}^*}{p_{\alpha\gamma}(\rho^{\mathrm{true}})} \\ a_{\alpha\gamma} &= C^T \mathbf{vec}\; O_{\alpha\gamma} \in \mathbf{C}^{n^2} \end{aligned}$$

and $C \in \mathbf{R}^{n^2\times n^2-1}$ is part of the unitary matrix $W = [c\; C] \in \mathbf{R}^{n^2\times n^2}$ in the singular value decomposition,

$$\mathbf{vec}\; I_n = W \begin{bmatrix} \sqrt{n} \\ 0_{n^2-1} \end{bmatrix}, \in \mathbf{R}^{n^2} \tag{16}$$

In general it is difficult to determine if any estimate will achieve the Cramér-Rao lower bound. Under the conditions stated, the ML estimate, $\rho^{\mathrm{ML}}(\ell)$, the solution to (13), approaches ρ^{true} with probability one, asymptotically as ℓ_{expt} increases, and the asymptotic distribution becomes Gaussian with covariance given by the Cramér-Rao bound.

The experiment design problem can be expressed by the following optimization problem in the vector of integers ℓ:

$$\begin{array}{l} \text{minimize } \; V(\ell,\rho^{\mathrm{true}}) \\ \text{subject to } \; \displaystyle\sum_{\gamma} \ell_\gamma = \ell_{\mathrm{expt}} \\ \qquad\qquad \text{integer } \ell_\gamma \geq 0 \end{array} \tag{17}$$

where ℓ_{expt} is the desired number of total experiments. The good news is that the objective, $V(\ell,\rho^{\mathrm{true}})$, is convex in ℓ (Boyd and Vandenburgh 2004, §7.5). Unfortunately, there are still two impediments: (i) restricting ℓ to a vector of integers makes the problem

[4] Proofs are eliminated to save space and are available upon request.

combinatorial; (ii) the lower-bound function $V(\ell, \rho^{\mathrm{true}})$ depends on the true value, ρ^{true}. All of these difficulties can be alleviated to some extent. For (i) we can use the convex relaxation described in (Boyd and Vandenburgh 2004, §7,5). For (ii) we can solve the relaxed experiment design problem with either a set of "what-if" estimates as surrogates for ρ^{true}, or use nominal values to start and then "bootstrap" to more precise values by iterating between state estimation and experiment design. We now explain how to perform these steps.

Relaxed experiment design for state estimation Following the procedure in (Boyd and Vandenburgh 2004), introduce the variables $\lambda_\gamma = \ell_\gamma/\ell_{\mathrm{expt}}$, each of which is the fraction of the total number of experiments performed in configuration γ. Since all the ℓ_γ and ℓ_{expt} are non-negative integers, each λ_γ is non-negative and *rational*, specifically an integer multiple of $1/\ell_{\mathrm{expt}}$, and in addition, $\sum_\gamma \lambda_\gamma = 1$. Let $\widehat{\rho}$ denote a surrogate for ρ^{true}, *e.g.*, an estimate or candidate value of ρ^{true}. Using (15) gives,

$$V(\ell = \ell_{\mathrm{expt}}\lambda, \widehat{\rho}) = \frac{1}{\ell_{\mathrm{expt}}} V(\lambda, \widehat{\rho}) \tag{18}$$

and,

$$\begin{aligned} V(\lambda, \widehat{\rho}) &= \mathrm{Tr}\, G(\lambda, \widehat{\rho})^{-1} \\ G(\lambda, \widehat{\rho}) &= \sum_\gamma \lambda_\gamma G_\gamma(\widehat{\rho}) \end{aligned} \tag{19}$$

Hence, the objective function $V(\ell, \widehat{\rho})$ can be replaced with $V(\lambda, \widehat{\rho})$ and the experiment design problem (17) is equivalent to.

$$\begin{aligned} &\text{minimize } V(\lambda, \widehat{\rho}) \\ &\text{subject to } \sum_\gamma \lambda_\gamma = 1,\ \lambda_\gamma \geq 0 \\ &\qquad \lambda_\gamma \text{ – integer multiple of } 1/\ell_{\mathrm{expt}} \end{aligned} \tag{20}$$

The objective is now a convex function of the λ_γ, but it is is still a combinatorial problem because the λ_γ are constrained to each be an integer multiple of $1/\ell_{\mathrm{expt}}$.

If λ_γ is only otherwise constrained to the non-negative reals, then this has the effect of relaxing the constraint that the ℓ_γ are integers. As phrased in (Boyd and Vandenburgh 2004), the *relaxed* experiment design problem is:

$$\begin{aligned} &\text{minimize } V(\lambda, \widehat{\rho}) = \mathrm{Tr}\big(\sum_\gamma \lambda_\gamma G_\gamma(\widehat{\rho})\big)^{-1} \\ &\text{subject to } \sum_\gamma \lambda_\gamma = 1 \\ &\qquad \lambda_\gamma \geq 0 \end{aligned} \tag{21}$$

The objective is convex, the equality constraint is linear, and the inequality constrains are convex, hence, this is a convex optimization problem in $\lambda \in \mathbf{R}^{n_{\mathrm{cfg}}}$. Let λ^{opt} denote the optimal solution to (21). Since the problem no longer depends on ℓ_{expt}, λ^{opt} can be viewed as a distribution of experiments per configuration.[5] There is also no guaranty that $\ell_{\mathrm{expt}}\lambda^{\mathrm{opt}}$ is a vector of integer multiples of $1/\ell_{\mathrm{expt}}$. A practical choice for obtaining a vector of integer multiples of $1/\ell_{\mathrm{expt}}$ is,

$$\ell_{\mathrm{expt}}^{\mathrm{round}} = \mathbf{round}\left\{\ell_{\mathrm{expt}}\lambda^{\mathrm{opt}}\right\} \tag{22}$$

If ℓ^{opt} is the (unknown) integer vector solution to (17), then we have the relations:

$$V(\ell_{\mathrm{expt}}^{\mathrm{round}}, \widehat{\rho}) \geq V(\ell^{\mathrm{opt}}, \widehat{\rho}) \geq V(\ell_{\mathrm{expt}}\lambda^{\mathrm{opt}}, \widehat{\rho}) \tag{23}$$

The gap within which falls the optimal solution can be no worse than the difference between $V(\ell_{\mathrm{expt}}^{\mathrm{round}}, \widehat{\rho})$ and $V(\ell_{\mathrm{expt}}\lambda^{\mathrm{opt}}, \widehat{\rho})$, which can be computed solely from λ^{opt}. If the gap is sufficiently small then for all practical purposes the "optimal" solution is λ^{opt}. From now on we will refer to λ^{opt} as the optimal solution rather than the relaxed optimal.

Bootstraping A standard approach used to circumvent not knowing the true state needed to optimize the experiment design is to procede adaptively, or by "bootstraping." The idea is to use the current estimate of the initial state found from (13), then solve (21), and then repeat. The algorithm at the k-th iteration looks like this:

$$\begin{aligned} \widehat{\rho}(k) &= \arg\min_\rho V(\widehat{\ell}(k-1), \rho) \\ \lambda^{\mathrm{opt}}(k) &= \arg\min_\lambda V(\lambda, \rho = \widehat{\rho}(k)) \\ \widehat{\ell}(k) &= \mathbf{round}\left\{\ell_{\mathrm{expt}}\lambda^{\mathrm{opt}}(k)\right\} \end{aligned} \tag{24}$$

The initial distribution $\widehat{\ell}(0)$ could be chosen as uniform, *e.g.*, the same for a not too large number of configurations. The algorithm could also start by first solving for a distribution from an initial state surrogate. In each iteration we could also vary ℓ_{expt}. Although each optimization is convex, the joint problem may not be. Conditions for convergence would need to be investigated as well as establishing that this method is efficient, *i.e.*, reduces the number of trials. We will not pursue this any further here.

2.4 *Example: Experiment Design for State Estimation*

A schematic of an apparatus for state tomography of a photon specified by the quantum state (density matrix) ρ is shown in figure 1.

---

[5] **Caveat emptor**: The relaxed optimal experiment design distribution, λ^{opt}, is optimal with respect to the initial state $\widehat{\rho}$, a surrogate for ρ^{true}. Thus, λ^{opt} is *not* optimal with respect to ρ^{true}. This should be no surprise because the underlying goal is to find a good estimate of ρ^{true}.

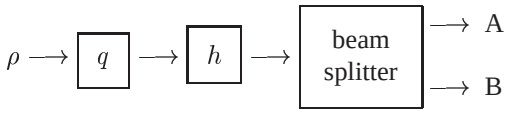

Fig. 1. Detection apparatus for single-photon tomography

The set up has two photon-counting detectors, A, B. There are two continuous variable settings for the quarter-wave plate and half-wave plate angles q, h. The objective is to determine the optimal settings of these parameters and number of experiments per setting for estimation of the state ρ using as data the photon counts from the two detectors. For these type of optical experiments there is significant cost (in time) associated with changing wave plate angles and very little cost (in time) for an experiment. As a result, although a uniform distribution can produce a good quality estimate, it is a *very* costly in terms of changing wave plate angles.

Suppose the wave plate settings are,

$$\{\, h_\gamma,\ q_\gamma \mid \gamma = 1, \ldots, n_{\rm cfg} \,\} \tag{25}$$

Assume also that the incoming state *always* is one photon, never none. Hence, $\rho \in \mathbf{C}^{2\times 2}$. Assuming each detector has efficiency $\eta \leq 1$ and a non-zero dark count probability δ, then there are four possible outcomes at detectors A,B denoted by the outcome indices $\alpha \in \{10,\ 01,\ 00,\ 11\}$. Following (Grice and Walmsley 1996, Walmsley and Waxer 1998) the probability of a dark count is denoted by the conditional probability,

$$\nu_{1|0} = \delta \tag{26}$$

where $1|0$ means the detector has fired "1" given that no photon is present at the detector "0." As shown in (Grice and Walmsley 1996), it therefore follows that the probability that the detector does not fire "0" although a photon is present at the detector "1" is given by,

$$\nu_{0|1} = (1-\eta)(1-\delta) \tag{27}$$

Here $1-\eta$ is the probability of no detection and $1-\delta$ is the probability of no dark count. The remaining conditional probabilities are, by definition, constrained to obey:

$$\begin{aligned} \nu_{1|0} + \nu_{0|0} &= 1 \\ \nu_{1|1} + \nu_{0|1} &= 1 \end{aligned} \tag{28}$$

The probabilities for the firing patterns in are thus given by (5) with the following observables $M_{\alpha\gamma}$:

$$\begin{aligned} M_{10,\gamma} &= \nu_{1|1}\nu_{0|0}M_\gamma^{10} + \nu_{1|0}\nu_{0|1}M_\gamma^{01} \\ M_{01,\gamma} &= \nu_{0|1}\nu_{1|0}M_\gamma^{10} + \nu_{0|0}\nu_{1|1}M_\gamma^{01} \\ M_{00,\gamma} &= \nu_{0|1}\nu_{0|0}M_\gamma^{10} + \nu_{0|0}\nu_{0|1}M_\gamma^{01} \\ M_{11,\gamma} &= \nu_{1|1}\nu_{1|0}M_\gamma^{10} + \nu_{1|0}\nu_{1|1}M_\gamma^{01} \end{aligned} \tag{29}$$

where $\left\{M_\gamma^\alpha\right\}$ are the noise-free observables. Simulations were performed for two input state cases:

$$\begin{aligned} &\textbf{pure state:} \quad \rho_{\rm pure} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\ &\textbf{mixed state:} \ \rho_{\rm mixd} = \begin{bmatrix} 0.6 & -0.2i \\ 0.2i & 0.4 \end{bmatrix} \end{aligned} \tag{30}$$

For each input state case we computed $\lambda^{\rm opt}$ with and without "noise:"

$$\begin{aligned} \text{no noise} &\begin{cases} \text{detector efficiency} & \eta = 1 \\ \text{dark count probability} & \delta = 0 \end{cases} \\ \text{yes noise} &\begin{cases} \text{detector efficiency} & \eta = 0.75 \\ \text{dark count probability} & \delta = 0.05 \end{cases} \end{aligned}$$

For all cases and noise conditions we used the wave plate settings:

$$\begin{aligned} h_i &= (i-1)(5^\circ),\ i = 1, \ldots, 10 \\ q_i &= (i-1)(5^\circ),\ i = 1, \ldots, 10 \end{aligned} \tag{31}$$

Both angles are set from 0 to 45° in 5° increments. This yields a total of $n_{\rm cfg} = 10^2 = 100$ configurations corresponding to all the wave plate combinations. The optmal distribution, $\lambda^{\rm opt}$, was obtained by solving (21) using a Barrier method (Boyd and Vandenburgh 2004, Ch.10).

Figure 2 shows the optimal distributions $\lambda^{\rm opt}$ versus configurations $\gamma = 1, \ldots, 100$ for all four test cases: two input states with and without noise. Observe that the optimal distributions are *not* uniform but are concentrated near the same particular wave plate settings. These settings are very close to those established in (James *et al.* 2001).

To check the gap between the relaxed optimum $\lambda^{\rm opt}$ and the unknown integer optimum we appeal to (22)-(23). Table 1 confirms that these distributions are good approximation to the unknown optimal integer solution for even not so large $\ell_{\rm expt}$ for the two state cases with no noise.

Table 1. Gap between sub-optimal and relaxed solutions.

$\ell_{\rm expt}$	$\frac{V(\ell_{\rm expt}\lambda^{\rm opt}_{\rm pure},\rho_{\rm pure})}{V(\ell^{\rm round}_{\rm expt}(\rho_{\rm pure}),\rho_{\rm pure})}$	$\frac{V(\ell_{\rm expt}\lambda^{\rm opt}_{\rm mixd},\rho_{\rm mixd})}{V(\ell^{\rm round}_{\rm expt}(\rho_{\rm mixd}),\rho_{\rm mixd})}$
100	.9797	.7761
1000	.9950	.9735
10000	.9989	.9954

Table 2 compares the distributions for optimal and uniform distributions at all 100 angles by examining the minimum number of experiments required to obtain an RMS estimation error of no more than 0.01.

Table 2. Comparison of optimal and uniform distributions.

input state	optimal $\mathbf{round}\left\{\ell_{\mathrm{expt}}\lambda^{\mathrm{opt}}\right\}$ $n_{\mathrm{cfg}} = 100$	uniform $(1/100)\,\mathbf{1}(\lambda^{\mathrm{opt}})$ $n_{\mathrm{cfg}} = 100$
ρ_{pure}, no noise	20308	29274
ρ_{pure}, yes noise	37775	52825
ρ_{mixd}, no noise	41890	64780
ρ_{mixd}, yes noise	61049	94385

Although the uniform distribution at all 100 angles does not require a significant increase in the number of experiments, as already mentioned it is a *very* costly approach in terms of the time required to change the wave plate angles. From figure 2 it is clear that many of the wave-plate angles can be eliminated thereby leading a suboptimal experiment design with only a few settings. Although not shown, these suboptimal settings do not require a significant increase in the number of experiments required to achieve the desired estimation accuracy of 0.01.

3. CONCLUDING REMARKS

The results presented show that an efficient numerical method based on convex programming can optimize the experiment for quantum state tomography. In addition, the estimation of the state and/or process using data from non-continuing measurements is copasetic with Maximum Likelihood Estimation. Both the experiment design and estimation work naturally together and both can be solved using convex optimization methods. The same methods can be used for quantum process tomography and Hamiltonian parameter estimation (Kosut *et al.* 2005).

4. REFERENCES

S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004;

H. Cramér, *Mathematical Methods of Statistics*, Princeton Press, 1946.

W. P. Grice and I. A. Walmsley, "Homodyne detection in a photon counting application," *Journal of Modern Optics*, 43, 795, (1996).

D. F. V. James, P. G. Kwiat, W. J. Munroe, and A. G. White, "Measurement of qubits," *Phys. Rev. A*, vol. 64 052312, 2001.

R. L. Kosut, H. Rabitz, and I. A. Walmsley, "Maximum likelihood identification of quantum systems for control design," *13th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, 27-29 August, 2003.

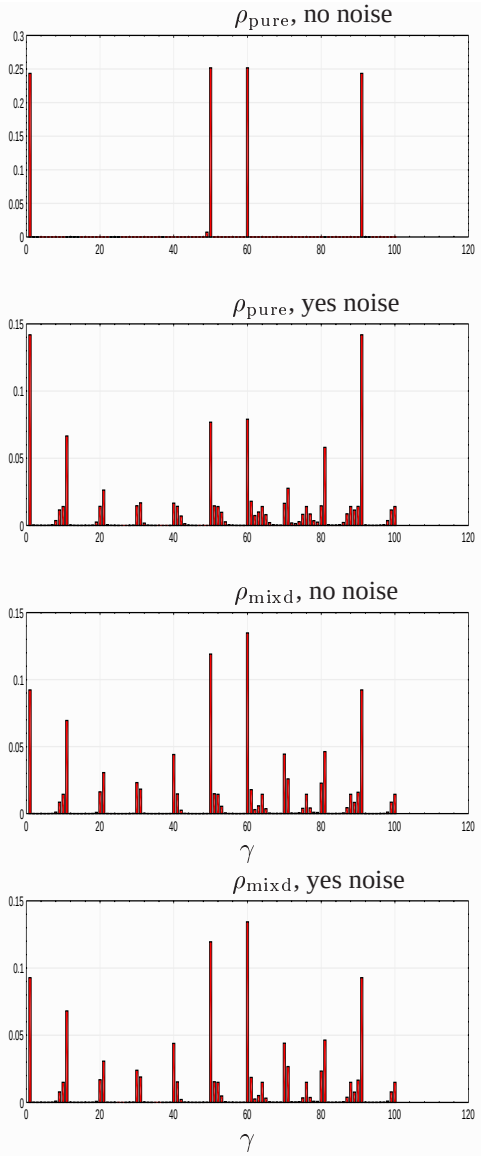

Fig. 2. Optimal distributions

R. L. Kosut, I. A. Walmsley and H. Rabitz, "Optimal Experiment Design for Quantum State and Process Tomography and Hamiltonian Parameter Estimation," in preparation.

M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge, 2000.

M. G. A. Paris, G. M. D'Ariano, and M. F. Sacchi, "Maximum likelihood method in quantum estimation," arXiv: quant-ph/ 0101071 v1, 16 Jan 2001.

M.Q. Phan and H. Rabitz, "Learning control of quantum-mechanical systems by laboratory identification of effective input-output maps," *Chem. Phys.*, 217, 389-400 (1997).

F. Verstraete, A. C. Doherty and H. Mabuchi, "Sensitivity optimization in quantum parameter estimation," *Phys. Rev. A* 64, 032111 (2001), quant-ph/0104116 v1, 24 Apr 2001.

I. A. Walmsley and L. Waxer, "Emission tomography for quantum state measurement in matter," *J. Phys. B: At. Mol. Opt. Phys.*, **31** (1998) 1825-1863.