Optimizing quantum process tomography with unitary **2**-designs

2008 J. Phys. A: Math. Theor. 41 055308

(http://iopscience.iop.org/1751-8121/41/5/055308)

IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **41** (2008) 055308 (26pp) doi:10.1088/1751-8113/41/5/055308

Optimizing quantum process tomography with unitary 2-designs

A J Scott

Centre for Quantum Dynamics, Griffith University, Brisbane, Queensland 4111, Australia
and
Centre for Quantum Computer Technology, Griffith University, Brisbane, Queensland 4111, Australia

E-mail: andrew.scott@griffith.edu.au

Received 8 November 2007, in final form 6 December 2007
Published 23 January 2008
Online at stacks.iop.org/JPhysA/41/055308

Abstract

We show that weighted unitary 2-designs define optimal measurements on the system-ancilla output state for ancilla-assisted process tomography of unital quantum channels. Examples include complete sets of mutually unbiased unitary-operator bases. Each of these specifies a minimal series of optimal orthogonal measurements. General quantum channels are also considered.

PACS numbers: 03.65.Wj, 03.67.−a, 02.10.Ox

1. Introduction

Fundamental to the fabrication of quantum information processing devices [1], such as quantum teleporters, key distributors, cloners, gates, and indeed, quantum computers, is the ability to precisely determine an unknown transformation on a quantum system. Quality assurance requires a complete characterization of these devices, which can be accomplished through a procedure known as *quantum process tomography* [2]: for judicious choices of initial system states, the transformation is uniquely identified by the outcomes of measurements on the transformed states.

The approach of *ancilla-assisted* quantum process tomography [3, 4] is to encode all information about the transformation into a single bipartite system-ancilla quantum state, and thus completely reduce the problem to that of quantum state tomography [2]. A sequence of measurements on identically prepared copies of this state will then reveal the particular transformation under examination. It is known that, for linear tomographic reconstructions of general quantum states, the most robust measurements against statistical error are described by tight informationally complete positive-operator-valued measures (tight IC-POVMs) [5]. Such measures derive their name from a related concept in frame theory, called a tight frame [6], and are equivalent to weighted complex projective 2-designs [5, 7–9].

The nonselective evolution of an open quantum system is described by a completely positive, trace preserving, linear transformation on quantum states. Such transformations are called *quantum channels* within the context of quantum information theory [1], as they also describe the degradation of information encoded in quantum states after transmission through a noisy communication channel. A *unital* quantum channel is one which fixes the maximally mixed state. These include all probabilistic applications of unitary operators, and thus, within the context of process tomography, form the relevant subclass of channels describing closed-system quantum dynamics.

In this paper, we study ancilla-assisted process tomography of general and unital quantum channels. The possible system-ancilla output states of relevance are then naturally housed in proper convex subsets of the set of all quantum states, and thus permit optimizations of the measurement over that necessary to identify a general quantum state. We find that the most robust measurements against statistical error, when they exist, are again described by tight POVMs, though a generalization thereof. In the unital case, these POVMs are equivalent to *weighted unitary 2-designs* [10, 11], but in the general case, they are shown not to exist.

This paper is organized as follows. Sections 2 and 3 review quantum process and state tomography, respectively, paying particular attention to the pertinent case of ancilla-assisted process tomography of quantum channels. Section 4 generalizes results of [5], characterizing the structure of POVMs that are optimal for linear quantum state tomography when a member of a convex subset of all possible quantum states need only be distinguished from other members. In section 5, we introduce the concept of a weighted unitary t-design, reviewing known results and presenting new ones. Finally, in section 6 we make the connection between weighted unitary 2-designs and the POVMs that optimize ancilla-assisted process tomography of unital quantum channels. The paper then concludes in section 7 where open problems are discussed. In addition, an appendix sets the superoperator notation used throughout this paper by reviewing a general class of transformations on quantum systems called quantum operations.

2. Quantum process tomography

The purpose of this paper is to optimize the measurements used for ancilla-assisted process tomography of quantum channels, and in particular, unital quantum channels. Quantum channels are nonselective quantum operations. The appendix provides some background to this broad class of transformations on quantum systems and introduces important concepts and notations relevant to the current study of channels. We will proceed by first describing process tomography for unitary operations. This will lead naturally into that for channels.

The dynamical evolution of a closed quantum system $\mathcal{H}_\mathrm{s} = \mathbb{C}^d$ is described by a unitary operator $U \in \mathrm{U}(d)$. In the absence of any physical description of the system, we expect that the Haar probability measure μ on $\mathrm{U}(d)$ most accurately reflects our state of knowledge of U. One method to determine U is then to couple $\mathcal{H}_\mathrm{s}$ to an auxiliary system $\mathcal{H}_\mathrm{a} = \mathbb{C}^{d_\mathrm{a}}$ (called the *ancilla*) and allow the combined system $\mathcal{H}_\mathrm{s} \otimes \mathcal{H}_\mathrm{a}$ to evolve from some initially known state, ρ_i say, to

$$\rho = (U \otimes I)\rho_\mathrm{i}(U^\dagger \otimes I). \tag{2.1}$$

A measurement on the combined system will then provide information on U. By repeating this procedure many times over, perhaps on different input states, U can be determined completely. This method of determining quantum dynamics is called *quantum process tomography*. Although the ancilla could be removed if the initial state were varied, in this

paper we investigate the opposite extreme by choosing $d_a = d$ and then $\rho_i = |I\rangle\langle I|$, fixed, where for any $V \in \mathrm{U}(d)$ we define

$$|V\rangle = (V \otimes I)|I\rangle := \frac{1}{\sqrt{d}} \sum_k V|k\rangle \otimes |k\rangle. \tag{2.2}$$

The pure state $|V\rangle\langle V|$ is a maximally entangled state of $\mathcal{H}_s \otimes \mathcal{H}_a$, and in fact, all maximally entangled states can be written in this form. The output state is $\rho = |U\rangle\langle U|$. Note that U can be found from $|U\rangle$ (and vice versa) through the relation $\langle j|U|k\rangle = \sqrt{d}(\langle j| \otimes \langle k|)|U\rangle$, a special case of the Jamiołkowski isomorphism below.

The determination of an unknown unitary U is thus equivalent to the determination of an unknown maximally entangled state $|U\rangle$. The latter can be accomplished through quantum state tomography. It is unrealistic, however, to presume that each system evolution in the above tomographic procedure can be performed identically. The class of quantum states under examination should thus be broadened to include any classical mixture of maximally entangled states: $\rho = \sum_k r_k |U_k\rangle\langle U_k|$, where each $r_k > 0$ and $\sum_k r_k = 1$. This is the output state of a quantum channel.

The (nonselective) evolution of an open quantum system is described by a *quantum channel*, i.e., a superoperator $\mathcal{E} \in \mathrm{End}(\mathrm{End}(\mathbb{C}^d))$ which is both trace preserving and completely positive (see the appendix). The channel is said to be *unital* if it fixes the maximally mixed state: $\mathcal{E}(I) = I$. Unital channels include all unitary operations $U \odot U^\dagger$, and moreover, all random-unitary channels, i.e., those which can be implemented by probabilistic applications of unitary operators (as above): $\mathcal{E} = \sum_k r_k U_k \odot U_k{}^\dagger$. In dimensions $d \geqslant 3$, however, there exist unital channels which cannot be decomposed in this way [12]. Although random-unitary channels are those channels which are most relevant to the study of closed-system dynamics, within this context, we will consider the entire class of unital channels together.

The process tomography of a quantum channel follows that for a unitary operator. The output state corresponding to the input $\rho_i = |I\rangle\langle I|$ completely determines the channel:

$$\rho = (\mathcal{E} \otimes \mathcal{I})(\rho_i) \quad \leftrightarrow \quad \mathcal{E} = d \sum_{j,k,l,m} \mathrm{tr}[(|m\rangle\langle j| \otimes |l\rangle\langle k|)\rho]|j\rangle\langle k| \odot |l\rangle\langle m|. \tag{2.3}$$

This is the so-called *Jamiołkowski isomorphism* [13]. It is important that the basis used in the right-hand side (RHS) of equation (2.3) is that in the definition of $|I\rangle\langle I|$. Note that

$$\mathrm{tr}_s(\rho) = \frac{1}{d} \sum_{j,k} \mathrm{tr}[\mathcal{E}(|j\rangle\langle k|)]|j\rangle\langle k| = \frac{1}{d} \sum_{j,k} \mathrm{tr}[|j\rangle\langle k|]|j\rangle\langle k| = \frac{1}{d} I, \tag{2.4}$$

since all quantum channels are trace preserving, and additionally,

$$\mathrm{tr}_a(\rho) = \frac{1}{d} \sum_{j,k} \mathcal{E}(|j\rangle\langle k|)\mathrm{tr}[|j\rangle\langle k|] = \frac{1}{d}\mathcal{E}(I) = \frac{1}{d} I, \tag{2.5}$$

for unital channels. This means that the classes of output states of quantum channels, with fixed input $\rho_i = |I\rangle\langle I|$, do not include all types of quantum states. The same could not be said if $\mathcal{E}$ were trace decreasing, i.e., a path of a general quantum operation. Denote by

$$\mathrm{Q}(\mathcal{H}) := \{A \in \mathrm{End}(\mathcal{H}) | A \geqslant 0, \mathrm{tr}(A) = 1\} \tag{2.6}$$

the set of all quantum states for $\mathcal{H}$. The two convex subsets of $\mathrm{Q}(\mathcal{H}_s \otimes \mathcal{H}_a)$,

$$\mathscr{Q}^{\mathrm{gc}} := \{\rho \in \mathrm{Q}(\mathcal{H}_s \otimes \mathcal{H}_a) | \mathrm{tr}_s(\rho) = I/d\} \quad \text{and} \tag{2.7}$$

$$\mathscr{Q}^{\mathrm{uc}} := \{\rho \in \mathrm{Q}(\mathcal{H}_s \otimes \mathcal{H}_a) | \mathrm{tr}_s(\rho) = \mathrm{tr}_a(\rho) = I/d\}, \tag{2.8}$$

then correspond to the outputs of, respectively, general and unital quantum channels.

Let $\{\lambda_k\}_{k=0}^{d^2-1}$ be an orthonormal Hermitian operator basis for $\mathrm{End}(\mathbb{C}^d)$ with the choice $\lambda_0 = I/\sqrt{d}$. The remaining operators, $\lambda_1, \ldots, \lambda_{d^2-1}$, then span the (d^2-1)-dimensional subspace of traceless operators $\mathrm{tr}(\lambda_k) = \sqrt{d}\,\mathrm{tr}(\lambda_0\lambda_k) = 0$ for all $k > 0$. Every quantum state for $\mathcal{H} = \mathbb{C}^d \otimes \mathbb{C}^d$ is of course expressible in terms of this basis:

$$\rho = \sum_{j,k} r_{jk}\lambda_j \otimes \lambda_k. \tag{2.9}$$

The coefficients must be real, $r_{jk} \in \mathbb{R}$, but besides positivity of the state, the only other constraint is from normalization: $r_{00} = 1/d$. In contrast, the output state $\rho \in \mathscr{Q}^{\mathrm{gc}}$ has $r_{0k} = 0$ for all $k > 0$ and in the unital case, $\rho \in \mathscr{Q}^{\mathrm{uc}}$ has $r_{k0} = r_{0k} = 0$ for all $k > 0$. The number of outcomes of a measuring instrument capable of identifying one such output from within its class of output states can thus be reduced from d^4, the number necessary to identify a general quantum state, to $d^2(d^2-1)+1$ for general channels, or $(d^2-1)^2+1$ for unital channels. Ancilla-assisted process tomography of quantum channels is thus not equivalent to tomographic reconstructions of general system-ancilla quantum states.

We will conclude this section by describing how quantum states are naturally embedded in Euclidean space. This approach will later provide insight when we harness the concepts of frame theory. Embedded in the complex vector space $\mathrm{End}(\mathcal{H})$ is a real vector space of Hermitian operators

$$\mathrm{H}(\mathcal{H}) := \{A \in \mathrm{End}(\mathcal{H}) | A^\dagger = A\}. \tag{2.10}$$

Equipped with the Hilbert–Schmidt inner product inherited from $\mathrm{End}(\mathcal{H})$, $(A|B) := \mathrm{tr}(A^\dagger B)$, which induces the Frobenius norm, $\|A\| := \sqrt{(A|A)}$, the vector space $\mathrm{H}(\mathcal{H})$ forms a real Hilbert space: $\mathrm{H}(\mathbb{C}^D) \cong \mathbb{R}^{D^2}$. Within the context of ancilla-assisted process tomography it will be assumed that $\mathcal{H} = \mathcal{H}_\mathrm{s} \otimes \mathcal{H}_\mathrm{a} = \mathbb{C}^d \otimes \mathbb{C}^d \cong \mathbb{C}^D$, i.e., $D = d^2$. The above coefficients r_{jk} then define a canonical choice for the isomorphism to $\mathbb{R}^{D^2}$. Define the two subspaces

$$\mathrm{H}^{\mathrm{gc}} := \{A \in \mathrm{H}(\mathbb{C}^d \otimes \mathbb{C}^d) | \mathrm{tr}_1(A) = \mathrm{tr}(A)I/d\} < \mathrm{H}(\mathbb{C}^D), \quad \text{and,} \tag{2.11}$$

$$\mathrm{H}^{\mathrm{uc}} := \{A \in \mathrm{H}(\mathbb{C}^d \otimes \mathbb{C}^d) | \mathrm{tr}_1(A) = \mathrm{tr}_2(A) = \mathrm{tr}(A)I/d\} < \mathrm{H}^{\mathrm{gc}} < \mathrm{H}(\mathbb{C}^D), \tag{2.12}$$

which contain the convex sets $\mathscr{Q}^{\mathrm{gc}}$ and $\mathscr{Q}^{\mathrm{uc}}$, respectively, and have dimensions $d^2(d^2-1)+1$ and $(d^2-1)^2+1$.

In general, $\mathrm{Q}(\mathcal{H})$ is naturally embedded into the vector subspace of $\mathrm{H}(\mathcal{H})$ consisting of all traceless Hermitian operators

$$\mathrm{H}_0(\mathcal{H}) := \{A \in \mathrm{H}(\mathcal{H}) | \mathrm{tr}(A) = 0\} < \mathrm{H}(\mathcal{H}). \tag{2.13}$$

Define

$$\mathbf{\Pi}_0 := \mathbf{I} - \frac{1}{D}|I)(I| \tag{2.14}$$

which projects onto $\mathrm{H}_0(\mathcal{H})$. This projection defines an isometric embedding of $\mathrm{Q}(\mathbb{C}^D)$ into a (D^2-1)-dimensional real Hilbert space, $\mathrm{Q}(\mathbb{C}^D) \hookrightarrow \mathrm{H}_0(\mathbb{C}^D) \cong \mathbb{R}^{D^2-1}$,

$$|\rho_0) := \mathbf{\Pi}_0|\rho) = |\rho - I/D), \tag{2.15}$$

in which the images of pure states lie on a sphere, $\||\psi\rangle\langle\psi| - I/D\| = \sqrt{(D-1)/D}$, and the images of mixed states within. In the special case $D = 2$ the embedding is bijective into this sphere, realizing the Bloch-sphere representation of a qubit, but is otherwise only injective. By 'isometric' we mean that distances are preserved: $\|\rho_0 - \sigma_0\| = \|\rho - \sigma\|$.

It is important to recognize that both $\mathscr{Q}^{\mathrm{gc}}$ and $\mathscr{Q}^{\mathrm{uc}}$ are embedded into proper vector subspaces of $\mathrm{H}_0(\mathbb{C}^D)$ under $\mathbf{\Pi}_0$:

$$\mathscr{Q}^{\mathrm{gc}} \hookrightarrow \mathrm{H}_0^{\mathrm{gc}} := \mathbf{\Pi}_0 \mathrm{H}^{\mathrm{gc}} = \{A \in \mathrm{H}(\mathbb{C}^d \otimes \mathbb{C}^d) | \mathrm{tr}_1(A) = 0\}, \quad \text{and,} \tag{2.16}$$

$$\mathscr{Q}^{\mathrm{uc}} \hookrightarrow \mathrm{H}_0^{\mathrm{uc}} := \mathbf{\Pi}_0 \mathrm{H}^{\mathrm{uc}} = \{A \in \mathrm{H}(\mathbb{C}^d \otimes \mathbb{C}^d) | \mathrm{tr}_1(A) = \mathrm{tr}_2(A) = 0\}. \tag{2.17}$$

The dimensions of these subspaces are $d^2(d^2-1)$ and $(d^2-1)^2$, respectively. In section 4, we will show that POVMs corresponding to tight frames on these subspaces, if they exist, are uniquely optimal for ancilla-assisted process tomography.

3. Quantum state tomography

This section serves as an introduction to quantum state tomography and is adapted from [5, section 4]. Instead of using the complex vector space $\mathrm{End}(\mathbb{C}^D) \cong \mathbb{C}^{D^2}$ as a backdrop, however, we will use the embedded real vector space of Hermitian operators, $\mathrm{H}(\mathbb{C}^D) \cong \mathbb{R}^{D^2}$.

The outcome statistics of a quantum measurement on a system $\mathcal{H} = \mathbb{C}^D$ are described by a *positive-operator-valued measure (POVM)* [14]. That is, an operator-valued function defined on a σ-algebra over a set $\mathscr{X}$ of outcomes, $F : \mathfrak{B}(\mathscr{X}) \to \mathrm{H}(\mathcal{H})$, which satisfies (1) $F(\mathscr{E}) \geqslant 0$ for all $\mathscr{E} \in \mathfrak{B}(\mathscr{X})$ with equality if $\mathscr{E} = \emptyset$, (2) $F\left(\bigcup_{k=1}^{\infty} \mathscr{E}_k\right) = \sum_{k=1}^{\infty} F(\mathscr{E}_k)$ for any sequence of disjoint sets $\mathscr{E}_k \in \mathfrak{B}(\mathscr{X})$, and (3) the normalization constraint $F(\mathscr{X}) = I$. In this paper, we always take $\mathfrak{B}(\mathscr{X})$ to be the Borel σ-algebra. When a quantum measurement has a countable number of outcomes, the indexed set of POVM elements $\{F(x)\}_{x \in \mathscr{X}}$ completely characterizes F, and is thus often referred to as the 'POVM'. We will call such POVMs *discrete*.

We will need to express an arbitrary POVM F in a standard form. To do this, note that each POVM defines a natural scalar-valued *trace measure* $\tau(\mathscr{E}) := \mathrm{tr}[F(\mathscr{E})]$, which inherits the normalization $\tau(\mathscr{X}) = D$. Since each matrix element of F is a complex-valued measure which is absolutely continuous with respect to the non-negative finite measure τ, the POVM can be expressed as

$$F(\mathscr{E}) = \int_{\mathscr{E}} \mathrm{d}\tau(x) P(x), \tag{3.1}$$

where the *positive-operator-valued density (POVD)* $P : \mathscr{X} \to \mathrm{H}(\mathcal{H})$ is uniquely defined up to a set of zero τ-measure. The POVD P is of course the Radon–Nikodym derivative of F with respect to τ. Note that $\mathrm{tr}(P) = 1$, τ-almost everywhere. If P also has unit rank, τ-almost everywhere, then we call F a *rank-1* POVM. In the special case of a discrete POVM, $P(x) = F(x)/\tau(x)$.

An informationally complete POVM F [5, 15, 16] is one with the property that each quantum state ρ is uniquely determined by its measurement statistics, $p(\mathscr{E}) := \mathrm{tr}[F(\mathscr{E})\rho]$. A sequence of measurements on copies of a system in an unknown state, enabling an estimate of these statistics, will then reveal the state. This process is called *quantum state tomography*.

Definition 3.1. *A POVM $F : \mathfrak{B}(\mathscr{X}) \to \mathrm{H}(\mathcal{H})$ is called informationally complete with respect to $\mathscr{Q} \subseteq \mathrm{Q}(\mathcal{H})$ if for each pair of distinct quantum states $\rho \neq \sigma \in \mathscr{Q}$ there exists an event $\mathscr{E} \in \mathfrak{B}(\mathscr{X})$ such that $\mathrm{tr}[F(\mathscr{E})\rho] \neq \mathrm{tr}[F(\mathscr{E})\sigma]$. A POVM which is informationally complete with respect to $\mathrm{Q}(\mathcal{H})$ is called an informationally complete POVM (IC-POVM).*

For an arbitrary POVM F, define the Hermitian superoperator $\mathcal{F} : \mathrm{H}(\mathcal{H}) \to \mathrm{H}(\mathcal{H})$ by

$$\mathcal{F} := \int_{\mathscr{X}} \mathrm{d}\tau(x) |P(x))(P(x)|, \tag{3.2}$$

which is positive and bounded under the left–right action. The latter follows from the fact that $\mathrm{Tr}(\mathcal{F}) = \int_{\mathscr{X}} \mathrm{d}\tau(x)(P(x)|P(x)) \leqslant D$ for any POVM, with equality only for rank-1 POVMs. The image and coimage of a Hermitian superoperator are equal. We call this vector subspace

of $\mathrm{H}(\mathcal{H})$ the *support* of $\mathcal{F}$ and denote it by $\mathrm{supp}(\mathcal{F})$. Let $\mathrm{span}(\mathscr{Q})$ denote the subspace of $\mathrm{H}(\mathcal{H})$ spanned by members of $\mathscr{Q}$, and let $\mathrm{ri}(\mathscr{Q})$ denote the relative interior of $\mathscr{Q}$, which is the interior of $\mathscr{Q}$ as a subset of its affine hull. Now consider the following.

Proposition 3.2. *Let $F : \mathfrak{B}(\mathscr{X}) \to \mathrm{H}(\mathcal{H})$ be a POVM. Then F is informationally complete w.r.t. $\mathscr{Q} \subseteq \mathrm{Q}(\mathcal{H})$ if $\mathscr{Q} \subseteq \mathrm{supp}(\mathcal{F})$. Moreover, if $\mathrm{ri}(\mathscr{Q}) \neq \emptyset$, then F is informationally complete w.r.t. $\mathscr{Q}$ if and only if $\mathscr{Q} \subseteq \mathrm{supp}(\mathcal{F})$.*

Proof. Let $\mathscr{Q} \subseteq \mathrm{supp}(\mathcal{F})$. Then for the distinct quantum states $\rho \neq \sigma \in \mathscr{Q}$ we have

$$\int_{\mathscr{X}} \mathrm{d}\tau(x) |\mathrm{tr}[P(x)(\rho - \sigma)]|^2 = (\rho - \sigma | \mathcal{F} | \rho - \sigma) > 0, \tag{3.3}$$

since $\rho - \sigma \in \mathrm{supp}(\mathcal{F})$, being a vector subspace, and $\rho - \sigma \neq 0$. Thus there must exist an event $\mathscr{E} \in \mathfrak{B}(\mathscr{X})$ with

$$\int_{\mathscr{E}} \mathrm{d}\tau(x)\, \mathrm{tr}[P(x)(\rho - \sigma)] \neq 0, \tag{3.4}$$

or equivalently, $\mathrm{tr}[F(\mathscr{E})\rho] \neq \mathrm{tr}[F(\mathscr{E})\sigma]$. This means F is informationally complete w.r.t. $\mathscr{Q}$.

Now let F be informationally complete w.r.t. $\mathscr{Q}$, let $\mathrm{ri}(\mathscr{Q}) \neq \emptyset$, and suppose $\mathscr{Q} \nsubseteq \mathrm{supp}(\mathcal{F})$. There must then exist an operator $A \in \mathrm{span}(\mathscr{Q})$, $A \neq 0$, such that

$$(A|\mathcal{F}|A) = \int_{\mathscr{X}} \mathrm{d}\tau(x) |\mathrm{tr}[P(x)A]|^2 = 0, \tag{3.5}$$

which means $\mathrm{tr}(PA) = 0$, τ-almost everywhere. This operator is therefore traceless:

$$\mathrm{tr}(A) = \mathrm{tr}[F(\mathscr{X})A] = \int_{\mathscr{X}} \mathrm{d}\tau(x)\, \mathrm{tr}[P(x)A] = 0. \tag{3.6}$$

Now for any $\rho \in \mathrm{ri}(\mathscr{Q})$, if $\epsilon > 0$ is chosen small enough, then $\sigma = \rho + \epsilon A$ is also a member of $\mathscr{Q}$, and moreover, $\sigma \neq \rho$ with

$$\mathrm{tr}[F(\mathscr{E})\sigma] = \mathrm{tr}[F(\mathscr{E})\rho] + \epsilon \int_{\mathscr{E}} \mathrm{d}\tau(x)\, \mathrm{tr}[P(x)A] = \mathrm{tr}[F(\mathscr{E})\rho] \tag{3.7}$$

for all $\mathscr{E} \in \mathfrak{B}(\mathscr{X})$. Thus F could not have been informationally complete w.r.t. $\mathscr{Q}$. We must therefore have $\mathscr{Q} \subseteq \mathrm{supp}(\mathcal{F})$. □

This proposition is a straightforward but important observation. If a quantum state need only be distinguished from other members of a given convex subset $\mathscr{Q} \subseteq \mathrm{Q}(\mathcal{H})$, e.g. $\mathscr{Q}^{\mathrm{gc}}$ or $\mathscr{Q}^{\mathrm{uc}}$, then since the relative interior of any convex set is nonempty, proposition 3.2 enables us to focus on POVMs for which $\mathrm{supp}(\mathcal{F}) \supseteq \mathscr{Q}$. Equivalently, it enables us to focus on POVMs for which $\mathrm{supp}(\mathcal{F}) \geqslant \mathrm{span}(\mathscr{Q})$.

Note that any POVM F is informationally complete w.r.t. $\mathscr{Q} = \mathrm{supp}(\mathcal{F}) \cap \mathrm{Q}(\mathcal{H})$. With this choice there is a standard procedure for reconstructing a member, $\rho \in \mathscr{Q}$, in terms of its measurement outcome statistics, $p(\mathscr{E}) = \mathrm{tr}[F(\mathscr{E})\rho]$. Let $\tilde{\mathcal{F}}$ be the unique superoperator with $\mathrm{supp}(\tilde{\mathcal{F}}) = \mathrm{supp}(\mathcal{F})$ for which

$$\tilde{\mathcal{F}}\mathcal{F} = \mathcal{F}\tilde{\mathcal{F}} = \mathbf{\Pi}_{\mathcal{F}}, \tag{3.8}$$

where $\mathbf{\Pi}_{\mathcal{F}}$ denotes the projector onto $\mathrm{supp}(\mathcal{F})$. Of course, $\tilde{\mathcal{F}} = \mathcal{F}^{-1}$ when $\mathscr{Q} = \mathrm{Q}(\mathcal{H})$ (as in [5]). Now defining

$$|R) := \tilde{\mathcal{F}}|P), \tag{3.9}$$

the left–right action on $|\rho)$ of the identity

$$\int_{\mathscr{X}} \mathrm{d}\tau(x) |R(x))(P(x)| = \int_{\mathscr{X}} \mathrm{d}\tau(x) \tilde{\mathcal{F}}|P(x))(P(x)| = \tilde{\mathcal{F}}\mathcal{F} = \mathbf{\Pi}_{\mathcal{F}}, \tag{3.10}$$

allows state reconstruction in terms of the measurement statistics:

$$\rho = \int_{\mathscr{X}} \mathrm{d}\tau(x)\,\mathrm{tr}[P(x)\rho]R(x) = \int_{\mathscr{X}} \mathrm{tr}[\mathrm{d}F(x)\rho]R(x) = \int_{\mathscr{X}} \mathrm{d}p(x)R(x). \tag{3.11}$$

Although the reconstruction operator-valued density R is generally not positive, it inherits all other properties of P, i.e. $\int_{\mathscr{X}} \mathrm{d}\tau(x)R(x) = I$ and $\mathrm{tr}(R) = 1$ (see [5]). Finally, it is straightforward to confirm that

$$\tilde{\mathcal{F}} = \int_{\mathscr{X}} \mathrm{d}\tau(x)|R(x))(R(x)|. \tag{3.12}$$

Although we could now proceed directly to an analysis of optimality, we will first briefly show how to embed POVMs into Euclidean space, just as was done for quantum states. First note that, for an arbitrary POVM F, the subspace $\mathrm{H}_0(\mathcal{H}) < \mathrm{H}(\mathcal{H})$ is $\mathcal{F}$-invariant, and in fact,

$$\mathcal{F} = \mathbf{\Pi}_0\mathcal{F}\mathbf{\Pi}_0 + \frac{1}{D}|I)(I|, \tag{3.13}$$

since the identity operator is always a left–right eigenvector with unit eigenvalue:

$$\mathcal{F}|I) = \int_{\mathscr{X}} \mathrm{d}\tau(x)|P(x))(P(x)|I) = \int_{\mathscr{X}} \mathrm{d}\tau(x)|P(x)) = \int_{\mathscr{X}} |\mathrm{d}F(x)) = |I). \tag{3.14}$$

Although the projector $|I)(I|/D$ is fixed by the normalization of F, its complement is free,

$$\mathcal{F}_0 := \mathbf{\Pi}_0\mathcal{F}\mathbf{\Pi}_0 = \mathcal{F} - \frac{1}{D}|I)(I| = \int_{\mathscr{X}} \mathrm{d}\tau(x)|P_0(x))(P_0(x)|, \tag{3.15}$$

where we define

$$|P_0) := \mathbf{\Pi}_0|P) = |P - I/D). \tag{3.16}$$

It is the superoperator $\mathcal{F}_0$ which can be adjusted for optimality. The POVM is thus embedded into $\mathrm{H}_0(\mathcal{H})$,

$$|F_0(\mathscr{E})) := \mathbf{\Pi}_0|F(\mathscr{E})) = |F(\mathscr{E}) - \tau(\mathscr{E})I/D) = \int_{\mathscr{E}} \mathrm{d}\tau(x)|P_0(x)). \tag{3.17}$$

This means $\mathrm{tr}(F_0) = 0$ and $F_0(\mathscr{X}) = 0$. Note that our 'embedding' is anchored to the trace measure, however, in that we cannot find F from F_0 without knowledge of τ.

4. Optimal linear quantum state tomography

In this section, we decide which POVMs are the most robust against statistical error for linear tomographic reconstructions of quantum states. Our premise is that a member of some subset of all possible quantum states needs to be distinguished from other members. Although we will focus on the two convex subsets, $\mathscr{Q}^{\mathrm{gc}}$ and $\mathscr{Q}^{\mathrm{uc}}$, being those sets relevant to the process tomography of general and unital quantum channels, the following analysis will apply to other convex subsets $\mathscr{Q} \subseteq \mathrm{Q}(\mathcal{H})$ with similar properties. The choice $\mathscr{Q} = \mathrm{Q}(\mathcal{H})$ was considered in [5] and the following can be considered a generalization.

Let F be a POVM which is informationally complete w.r.t. $\mathscr{Q} \subseteq \mathrm{Q}(\mathcal{H})$. Throughout this section we assume a linear state-reconstruction formula valid for all $\rho \in \mathscr{Q}$ of the form

$$|\rho) = \int_{\mathscr{X}} \mathrm{d}p(x)|Q(x)) = \int_{\mathscr{X}} (\mathrm{d}F(x)|\rho)|Q(x)), \tag{4.1}$$

where $Q : \mathscr{X} \to \mathrm{span}(\mathscr{Q})$ is called a *reconstruction OVD*. By linearity, this formula implicitly presupposes that $\mathrm{supp}(\mathcal{F}) \geqslant \mathrm{span}(\mathscr{Q})$.

It is instructive to start with the special case of a discrete POVM, $\{F(x)\}_{x\in\mathscr{X}}$. Suppose that $y_1, \ldots, y_N$ are the outcomes of measurements on N identical copies of the state $\rho \in \mathscr{Q}$. One estimate for the outcome probabilities is then

$$\hat{p}(x) = \hat{p}(x; y_1, \ldots, y_N) := \frac{1}{N}\sum_{k=1}^{N} \delta(x, y_k), \tag{4.2}$$

which gives

$$\hat{\rho} = \hat{\rho}(y_1, \ldots, y_N) := \sum_{x\in\mathscr{X}} \hat{p}(x; y_1, \ldots, y_N) Q(x), \tag{4.3}$$

for an estimate of ρ. We will call $\hat{\rho}$ a *linear tomographic estimate* of ρ to distinguish it from more sophisticated choices, such as those from maximum likelihood estimation [17, 18] or Bayesian mean estimation [19–23].

The mean-squared Hilbert–Schmidt distance provides a measure of the expected error in our estimate,

$$e^{(F,Q)}(\rho) := \mathop{\mathrm{E}}_{y_1,\ldots,y_N} [\|\rho - \hat{\rho}(y_1, \ldots, y_N)\|^2] \tag{4.4}$$

$$= \sum_{x,y\in\mathscr{X}} \mathop{\mathrm{E}}_{y_1,\ldots,y_N} [(p(x) - \hat{p}(x))(p(y) - \hat{p}(y))](Q(x)|Q(y)) \tag{4.5}$$

$$= \frac{1}{N}\left(\sum_{x\in\mathscr{X}} p(x)(Q(x)|Q(x)) - \mathrm{tr}(\rho^2)\right) \tag{4.6}$$

$$=: \frac{1}{N}(\Delta_p(Q) - \mathrm{tr}(\rho^2)), \tag{4.7}$$

using equation (4.1) and given that

$$\mathop{\mathrm{E}}_{y_1,\ldots,y_N} [(p(x) - \hat{p}(x))(p(y) - \hat{p}(y))] = \frac{1}{N}(p(x)\delta(x, y) - p(x)p(y)), \tag{4.8}$$

which is an elementary calculation. Equation (4.7) is also a fitting description of the error for a POVM with a continuum of measurement outcomes if we define

$$\Delta_p(Q) := \int_{\mathscr{X}} \mathrm{d}p(x)(Q(x)|Q(x)) \tag{4.9}$$

in general. This is because a countable partition of the outcome set $\mathscr{X}$ allows any POVM to be approximated by a discrete POVM. Our estimate $\hat{p}$ remains a good approximation for the probability measure p, except now with x and $y_1, \ldots, y_N$ in equation (4.2) indicating members of the partition. In the limit of finer approximating partitions we again arrive at equation (4.7) for the average error, but now with equation (4.9) for $\Delta_p(Q)$. Since we have no control over the purity of ρ, it is this quantity which is now of interest.

The POVM which minimizes $\Delta_p(Q)$, and hence the error, will depend on the quantum state under examination. When $\mathscr{Q} = \mathrm{Q}(\mathbb{C}^D)$ it is natural to remove this dependence by averaging over all Hilbert-space orientations between the system and measurement apparatus. That is, we set $\rho = \rho(\sigma, U) := U\sigma U^\dagger$ where $\sigma \in \mathrm{Q}(\mathbb{C}^D)$ is fixed, and average $\Delta_p(Q)$ over random choices of $U \in \mathrm{U}(D)$. When $\mathscr{Q} = \mathscr{Q}^{\mathrm{gc}}, \mathscr{Q}^{\mathrm{uc}} \subseteq \mathrm{Q}(\mathcal{H}_\mathrm{s} \otimes \mathcal{H}_\mathrm{a})$ the natural procedure is to average over all local Hilbert-space orientations $U_\mathrm{s} \in \mathrm{U}(d)$ between the system and measurement apparatus, and all local Hilbert-space orientations $U_\mathrm{a} \in \mathrm{U}(d)$ between the

system and ancilla. The end result is the same, however. Setting $\rho = \rho(\sigma, U_s \otimes U_a) = (U_s \otimes U_a)\sigma(U_s \otimes U_a)^\dagger$ we take the average over all $U_s, U_a \in \mathrm{U}(d)$:

$$\int\!\!\int_{\mathrm{U}(d)} \mathrm{d}\mu(U_s)\,\mathrm{d}\mu(U_a)\Delta_p(Q) = \int\!\!\int_{\mathrm{U}(d)} \mathrm{d}\mu(U_s)\,\mathrm{d}\mu(U_a)$$

$$\times \int_{\mathscr{X}} \mathrm{tr}[\mathrm{d}F(x)(U_s \otimes U_a)\sigma(U_s \otimes U_a)^\dagger](Q(x)|Q(x)) \tag{4.10}$$

$$= \frac{1}{D}\int_{\mathscr{X}} \mathrm{tr}[\mathrm{d}F(x)]\mathrm{tr}(\sigma)(Q(x)|Q(x)) \tag{4.11}$$

$$= \frac{1}{D}\int_{\mathscr{X}} \mathrm{d}\tau(x)(Q(x)|Q(x)) \tag{4.12}$$

$$=: \frac{1}{D}\Delta_\tau(Q), \tag{4.13}$$

where μ is the unit Haar measure, using Shur's lemma for the integrals.

It would be presumptuous to take $\Delta_\tau(Q)$ as an error estimate for an arbitrary subset $\mathscr{Q} \subseteq \mathrm{Q}(\mathbb{C}^D)$ without further information on its structure. Nevertheless, assume that there is a natural set of possible 'orientations' $\mathscr{O} \subseteq \mathrm{U}(D)$ between $\mathcal{H} = \mathbb{C}^D$ and the measuring apparatus, and a probability measure ν on $\mathscr{O}$, with the property that for any $\sigma \in \mathscr{Q}$,

$$\int_{\mathscr{O}} \mathrm{d}\nu(U) U\sigma U^\dagger = \frac{1}{D} I. \tag{4.14}$$

Then $\int_{\mathscr{O}} \mathrm{d}\nu(U)\Delta_p(Q) = \Delta_\tau(Q)/D$ as above. We thus take $\mathscr{O} = \mathrm{U}(d) \otimes \mathrm{U}(d)$ and $\nu = \mu \times \mu$ when $\mathscr{Q} = \mathscr{Q}^{\mathrm{gc}}$ or $\mathscr{Q}^{\mathrm{uc}}$. Another example is

$$\mathscr{Q}^{\mathrm{cl}} := \{\rho \in \mathrm{Q}(\mathbb{C}^D) |\ \rho \text{ is diagonal in the standard basis}\}. \tag{4.15}$$

The members of $\mathscr{Q}^{\mathrm{cl}}$ might be described as 'classical' states, being convex combinations of basis states: $\rho = \sum_k r_k |k\rangle\langle k|$. Under random permutations U of basis elements, equation (4.14) is satisfied for any $\sigma \in \mathscr{Q}^{\mathrm{cl}}$. In general, we suspect that the above averaging of the error makes sense whenever $\mathscr{Q}$ is a convex subset of $\mathrm{Q}(\mathbb{C}^D)$, containing the completely mixed state I/D, and possessing a symmetry about this state described by equation (4.14).

We now proceed to an analysis of optimality. The above considerations are summarized as a definition,

$$e_{\mathrm{av}}^{(F,Q)}(\sigma) := \int_{\mathscr{O}} \mathrm{d}\nu(U)\, e^{(F,Q)}(\rho(\sigma, U)) = \frac{1}{ND}(\Delta_\tau(Q) - D\,\mathrm{tr}(\sigma^2)). \tag{4.16}$$

Our goal now is to find the optimal pairs (F, Q) which minimize $e_{\mathrm{av}}^{(F,Q)}$ for a given fixed $\mathscr{Q}$.

There are generally many different choices for the reconstruction OVD $Q : \mathscr{X} \to \mathrm{span}(\mathscr{Q})$ that satisfy equation (4.1). Our next task is to show that the canonical choice we encountered in section 3 is uniquely optimal. We thus minimize $\Delta_\tau(Q)$ over all Q while keeping F fixed. Our only constraint is that our state-reconstruction formula (equation (4.1)) remains valid for all $\rho \in \mathscr{Q}$. By linearity, this formula is then also valid for any member of $\mathrm{span}(\mathscr{Q})$, and therefore,

$$\int_{\mathscr{X}} |Q(x))(\mathrm{d}F(x)|\mathbf{\Pi}_{\mathscr{Q}} = \int_{\mathscr{X}} \mathrm{d}\tau(x)|Q(x))(P'(x)| = \mathbf{\Pi}_{\mathscr{Q}}, \tag{4.17}$$

where $\mathbf{\Pi}_{\mathscr{Q}}$ projects onto $\mathrm{span}(\mathscr{Q})$ and $|P') := \mathbf{\Pi}_{\mathscr{Q}}|P)$. This means $\{Q(x)\}_{x\in\mathscr{X}}$ is a dual frame to the frame $\{P'(x)\}_{x\in\mathscr{X}}$, w.r.t. τ, within the subspace $\mathrm{span}(\mathscr{Q})$. Consult Christensen

[24] for an introduction to frame theory (see also [5]). The following lemma shows that there is a unique canonical dual frame which is optimal (see [5, lemma 16] for a proof).

Lemma 4.1. *Let* $\{A(x)\}_{x\in\mathscr{X}}$ *be an operator frame w.r.t. the measure* α. *Then for all dual frames* $\{B(x)\}_{x\in\mathscr{X}}$,

$$\int_{\mathscr{X}} \mathrm{d}\alpha(x)(B(x)|B(x)) \geqslant \int_{\mathscr{X}} \mathrm{d}\alpha(x)(\tilde{A}(x)|\tilde{A}(x)), \tag{4.18}$$

with equality only if $B = \tilde{A}$, α*-almost everywhere, where* $\{\tilde{A}(x)\}_{x\in\mathscr{X}}$ *is the canonical dual frame, i.e.,* $|\tilde{A}) := \mathcal{A}^{-1}|A)$ *with* $\mathcal{A} := \int_{\mathscr{X}} \mathrm{d}\alpha(x)|A(x))(A(x)|$.

Let $\tilde{\mathcal{F}}'$ be the inverse of $\mathcal{F}' := \mathbf{\Pi}_{\mathscr{Q}}\mathcal{F}\mathbf{\Pi}_{\mathscr{Q}}$ in the subspace $\mathrm{span}(\mathscr{Q})$ and define $|R') := \tilde{\mathcal{F}}'|P')$. This means

$$\tilde{\mathcal{F}}' = \int_{\mathscr{X}} \mathrm{d}\tau(x)|R'(x))(R'(x)|. \tag{4.19}$$

Lemma 4.1 shows that

$$\Delta_\tau(Q) \geqslant \Delta_\tau(R') = \mathrm{Tr}(\tilde{\mathcal{F}}'), \tag{4.20}$$

with equality only if $Q = R'$, τ-almost everywhere. We thus make this choice and now minimize the quantity $\mathrm{Tr}(\tilde{\mathcal{F}}')$ over all POVMs. Define $\delta' := \dim \mathrm{span}(\mathscr{Q})$.

Lemma 4.2. *Let* $F : \mathfrak{B}(\mathscr{X}) \to \mathrm{H}(\mathbb{C}^D)$ *be a POVM which is informationally complete w.r.t.* $\mathscr{Q} \subseteq \mathrm{Q}(\mathbb{C}^D)$, *where* $\mathrm{ri}(\mathscr{Q}) \neq \emptyset$ *and* $I \in \mathrm{span}(\mathscr{Q})$. *Then*

$$\mathrm{Tr}(\tilde{\mathcal{F}}') \geqslant \frac{(\delta'-1)^2}{D-1} + 1, \tag{4.21}$$

with equality if and only if

$$\mathcal{F} = \frac{D-1}{\delta'-1}\mathbf{\Pi}_{\mathscr{Q}} + \frac{\delta'-D}{(\delta'-1)D}|I)(I|. \tag{4.22}$$

Proof. Since F is informationally complete w.r.t. $\mathscr{Q}$ and $\mathrm{ri}(\mathscr{Q}) \neq \emptyset$, by proposition 3.2, we must have $\mathrm{supp}(\mathcal{F}) \geqslant \mathrm{span}(\mathscr{Q})$. Thus $\mathcal{F}' = \mathbf{\Pi}_{\mathscr{Q}}\mathcal{F}\mathbf{\Pi}_{\mathscr{Q}}$ has δ' nonzero left–right eigenvalues: $\lambda_1, \ldots, \lambda_{\delta'} > 0$. One eigenvalue is fixed at unity, however, since $\mathcal{F}'|I) = \mathbf{\Pi}_{\mathscr{Q}}\mathcal{F}\mathbf{\Pi}_{\mathscr{Q}}|I) = |I)$ when $|I) \in \mathrm{span}(\mathscr{Q})$, given that $|I)$ is always an eigenvector of $\mathcal{F}$ (equation (3.14)). We thus take $\lambda_1 = 1$. The remaining eigenvalues satisfy

$$\sum_{k=2}^{\delta'} \lambda_k = \mathrm{Tr}(\mathcal{F}') - 1 \leqslant \mathrm{Tr}(\mathcal{F}) - 1 \leqslant D - 1, \tag{4.23}$$

given that $\mathrm{Tr}(\mathcal{F}') = \mathrm{Tr}(\mathbf{\Pi}_{\mathscr{Q}}\mathcal{F}) \leqslant \mathrm{Tr}(\mathcal{F})$, with equality if and only if $\mathrm{supp}(\mathcal{F}) = \mathrm{span}(\mathscr{Q})$, and $\mathrm{Tr}(\mathcal{F}) \leqslant D$, with equality if and only if F is a rank-1 POVM. Under this constraint, it is straightforward to show that

$$\mathrm{Tr}(\tilde{\mathcal{F}}') = \sum_{k=2}^{\delta'} \frac{1}{\lambda_k} + 1 \tag{4.24}$$

takes its minimum value if and only if $\lambda_2 = \cdots = \lambda_{\delta'} = (D-1)/(\delta'-1)$, or equivalently,

$$\mathcal{F}' = \frac{D-1}{\delta'-1}\left(\mathbf{\Pi}_{\mathscr{Q}} - \frac{1}{D}|I)(I|\right) + \frac{1}{D}|I)(I|. \tag{4.25}$$

Note that $\mathrm{Tr}(\mathcal{F}') = D$ with this choice, however, which requires $\mathrm{supp}(\mathcal{F}) = \mathrm{span}(\mathscr{Q})$. We must therefore have $\mathcal{F}' = \mathcal{F}$ and equation (4.25) is equivalent to equation (4.22). Finally,

the minimum value of equation (4.24) is $\mathrm{Tr}(\tilde{\mathcal{F}}') = (\delta' - 1) \cdot ((\delta' - 1)/(D - 1)) + 1 = (\delta' - 1)^2/(D - 1) + 1$. □

It is important to recognize that our condition for optimality (equation (4.22)) sets $\mathrm{supp}(\mathcal{F}) = \mathrm{span}(\mathscr{Q})$. We now proceed under this assumption, effectively replacing each primed symbol above by its unprimed counterpart. Furthermore, optimality requires a rank-1 POVM. This is because equation (4.22) gives $\mathrm{Tr}(\mathcal{F}) = D$, which is possible only for rank-1 POVMs. When $\mathscr{Q} = \mathrm{Q}(\mathbb{C}^D)$ we recover the tight rank-1 IC-POVMs described in [5]. Let us use the same terminology here.

Definition 4.3. *Let $F : \mathfrak{B}(\mathscr{X}) \to \mathrm{H}(\mathbb{C}^D)$ be a POVM. Then F is called tight if the OVD $\{P_0(x)\}_{x\in\mathscr{X}}$ forms a tight operator frame w.r.t. τ in* $\mathrm{supp}(\mathcal{F}_0)$*, i.e.,*

$$\mathcal{F}_0 := \int_{\mathscr{X}} \mathrm{d}\tau(x) |P_0(x))(P_0(x)| = a\mathbf{\Pi}_{\mathcal{F}_0}, \tag{4.26}$$

for some constant $a > 0$, or equivalently,

$$\mathcal{F} = a\mathbf{\Pi}_{\mathcal{F}} + \frac{1-a}{D}|I)(I|. \tag{4.27}$$

The constant satisfies $a \leqslant (D-1)/(\delta-1)$, where $\delta := \mathrm{rank}(\mathcal{F})$ (left–right rank), with equality only for rank-1 POVMs. Returning to equation (4.22) we see that tight rank-1 POVMs are precisely those which are optimal for linear quantum state tomography. That is,

$$\mathcal{F} = \frac{D-1}{\delta-1}\mathbf{\Pi}_{\mathcal{F}} + \frac{\delta - D}{(\delta-1)D}|I)(I| \tag{4.28}$$

if and only if F is a tight rank-1 POVM. The optimal state-reconstruction formula is given by equation (3.11), which now takes the form

$$\rho = \frac{\delta-1}{D-1}\int_{\mathscr{X}} \mathrm{d}p(x) P(x) - \frac{\delta - D}{D(D-1)} I, \tag{4.29}$$

since a straightforward calculation of $\tilde{\mathcal{F}}$ from equation (4.28) shows that

$$R = \frac{\delta-1}{D-1}P - \frac{\delta - D}{D(D-1)} I. \tag{4.30}$$

We now restate our findings in a theorem.

Theorem 4.4. *Let $F : \mathfrak{B}(\mathscr{X}) \to \mathrm{H}(\mathbb{C}^D)$ be a POVM which is informationally complete w.r.t. $\mathscr{Q} \subseteq \mathrm{Q}(\mathbb{C}^D)$, assumed convex and containing I/D, and let $\delta' = \dim \mathrm{span}(\mathscr{Q})$. Then for any fixed quantum state $\sigma \in \mathscr{Q}$,*

$$e_{\mathrm{av}}^{(F,Q)}(\sigma) \geqslant \frac{1}{ND}\left(\frac{(\delta'-1)^2}{D-1} + 1 - D\,\mathrm{tr}(\sigma^2)\right), \tag{4.31}$$

for all reconstruction OVDs Q. Furthermore, equality occurs if and only if $Q = R$ and F is a tight rank-1 POVM with $\mathrm{supp}(\mathcal{F}) = \mathrm{span}(\mathscr{Q})$.

Now consider the worst-case error. The average provides a lower bound:

$$e_{\mathrm{wc}}^{(F,Q)}(\sigma) := \sup_{U\in\mathscr{O}} e^{(F,Q)}(\rho(\sigma, U)) \tag{4.32}$$

$$\geqslant e_{\mathrm{av}}^{(F,Q)}(\sigma) \tag{4.33}$$

$$\geqslant \frac{1}{ND}\left(\frac{(\delta'-1)^2}{D-1} + 1 - D\,\mathrm{tr}(\sigma^2)\right). \tag{4.34}$$

Returning to equation (4.7), however, but now with $Q = R$ and $\rho = \rho(\sigma, U) = U\sigma U^\dagger$, we find

$$e^{(F,R)}(\rho(\sigma,U)) = \frac{1}{N}\left(\int_{\mathscr{X}} \mathrm{d}p(x)(R(x)|R(x)) - \mathrm{tr}(\sigma^2)\right) \tag{4.35}$$

$$= \frac{1}{N}\left(\frac{1}{D}\left(\frac{(\delta-1)^2}{D-1}+1\right)\int_{\mathscr{X}} \mathrm{d}p(x) - \mathrm{tr}(\sigma^2)\right) \tag{4.36}$$

$$= \frac{1}{ND}\left(\frac{(\delta-1)^2}{D-1}+1-D\,\mathrm{tr}(\sigma^2)\right), \tag{4.37}$$

when R satisfies equation (4.30) and P is rank 1, regardless of orientation $U \in \mathscr{O} \subseteq \mathrm{U}(D)$. Thus given $\delta = \delta'$ [$\mathrm{supp}(\mathcal{F}) = \mathrm{span}(\mathscr{Q})$] when $e_{\mathrm{av}}^{(F,Q)}$ is minimized, the following is a consequence.

Corollary 4.5. *Let $F : \mathfrak{B}(\mathscr{X}) \to \mathrm{H}(\mathbb{C}^D)$ be a POVM which is informationally complete w.r.t. $\mathscr{Q} \subseteq \mathrm{Q}(\mathbb{C}^D)$, assumed convex and containing I/D, and let $\delta' = \dim \mathrm{span}(\mathscr{Q})$. Then for any fixed quantum state $\sigma \in \mathscr{Q}$,*

$$e_{\mathrm{wc}}^{(F,Q)}(\sigma) \geqslant \frac{1}{ND}\left(\frac{(\delta'-1)^2}{D-1}+1-D\,\mathrm{tr}(\sigma^2)\right), \tag{4.38}$$

for all reconstruction OVDs Q. Furthermore, equality occurs if and only if $Q = R$ and F is a tight rank-1 POVM with $\mathrm{supp}(\mathcal{F}) = \mathrm{span}(\mathscr{Q})$.

Tight rank-1 POVMs are thus optimal for linear quantum state tomography in both an average and worst-case sense. In fact, they form the unique class of POVMs that achieve

$$e_{\mathrm{wc}}^{(F,R)}(\sigma) = e_{\mathrm{av}}^{(F,R)}(\sigma) = e^{(F,R)}(\rho(\sigma,U)) = \frac{1}{ND}\left(\frac{(\delta'-1)^2}{D-1}+1-D\,\mathrm{tr}(\sigma^2)\right). \tag{4.39}$$

The exact structure of these POVMs for $\mathscr{Q} = \mathscr{Q}^{\mathrm{gc}}$ and $\mathscr{Q} = \mathscr{Q}^{\mathrm{uc}}$, when they exist, will be explored in detail in section 6. To do so, however, we first need to explain the concept of a 'unitary t-design'. This is done in the following section. When $\mathscr{Q} = \mathrm{Q}(\mathbb{C}^D)$ we recover the results of [5, theorem 18 and corollary 19]. Lastly, consider $\mathscr{Q} = \mathscr{Q}^{\mathrm{cl}}$. Only $\delta' = D$ dimensions are then spanned, giving $\big(1 - \mathrm{tr}(\sigma^2)\big)/N$ for the minimum error. In particular, for pure states this is zero.

5. Unitary t-designs

The extension of spherical t-designs [7] to the unitary group was recently considered by Dankert *et al* [10] and Gross *et al* [11], and the following definition is equivalent to theirs. By a 'unitary t-design', however, we really mean a *projective* unitary t-design, in that each $\mathrm{e}^{\mathrm{i}\phi}U \in \mathrm{U}(d)$ should always be identified with U. With this in mind, let $U(x) \in \mathrm{U}(d)$ denote a representative from the equivalence class of unitaries $x \in \mathrm{PU}(d) = \mathrm{U}(d)/\mathrm{U}(1)$ and let μ denote the Haar measure on $\mathrm{PU}(d)$ with the normalization $\mu(\mathrm{PU}(d)) = 1$. A countable set $\mathscr{S}$ endowed with a weight function $w : \mathscr{S} \to (0, 1]$, where $\sum_{x\in\mathscr{S}} w(x) = 1$, will be called a *weighted set* and denoted by the pair $(\mathscr{S}, w)$.

Definition 5.1. *A finite weighted set $(\mathscr{D}, w)$, $\mathscr{D} \subset \mathrm{PU}(d)$, is called a weighted t-design (in dimension d) if*

$$\sum_{x\in\mathscr{D}} w(x)U(x)^{\otimes t} \otimes (U(x)^{\otimes t})^\dagger = \int_{\mathrm{PU}(d)} \mathrm{d}\mu(x)U(x)^{\otimes t} \otimes (U(x)^{\otimes t})^\dagger. \tag{5.1}$$

When $w(x) = 1/|\mathscr{D}|$ we recover the more common notion of an 'unweighted' t-design. Define $T := \sum_{j,k} |j\rangle\langle k| \otimes |k\rangle\langle j|$, which satisfies $\mathrm{tr}[(A \otimes B)T] = \mathrm{tr}(AB)$ and is called the *swap* (or *transposition*) since $T|\psi\rangle \otimes |\phi\rangle = |\phi\rangle \otimes |\psi\rangle$. By multiplying equation (5.1) on the right by $I^{\otimes(t-1)} \otimes T \otimes I^{\otimes(t-1)}$ and tracing out the inner pair of subsystems, we can immediately deduce that every weighted t-design is also a weighted $(t-1)$-design. Repeating this process t times shows that the normalization of w is in fact already implied by equation (5.1). Unweighted t-designs in $\mathrm{PU}(d)$ exist for every t and d:

Theorem 5.2 (Seymour and Zaslavsky [25])**.** *Let Ω be a path-connected topological space endowed with a measure ω that is finite and positive with full support and let $f : \Omega \to \mathbb{R}^m$ be a continuous, integrable function. Then there exists a finite set $\mathscr{X} \subseteq \Omega$ such that*

$$\frac{1}{|\mathscr{X}|} \sum_{x \in \mathscr{X}} f(x) = \frac{1}{\omega(\Omega)} \int_{\Omega} \mathrm{d}\omega(x) f(x). \tag{5.2}$$

The size of $\mathscr{X}$ may be any number, with a finite number of exceptions.

Corollary 5.3. *For each pair of positive integers t and d, and for all sufficiently large n, there exist (unweighted) unitary t-designs in dimension d of size n.*

Proof. Simply let $\Omega = \mathrm{PU}(d)$, $\omega = \mu$, and apply theorem 5.2 to

$$f(x) := U(x)^{\otimes t} \otimes (U(x)^{\otimes t})^{\dagger}, \tag{5.3}$$

which maps $\mathrm{PU}(d)$ into $\mathrm{End}(\mathbb{C}^d)^{\otimes 2t} \cong \mathbb{R}^{2d^{4t}}$. □

The task of finding t-designs is facilitated by the following theorem. Define the positive constant

$$\gamma(t, d) := \int_{\mathrm{PU}(d)} \mathrm{d}\mu(x) |\mathrm{tr}[U(x)]|^{2t}. \tag{5.4}$$

Theorem 5.4. *For any finite weighted set $(\mathscr{S}, w)$, $\mathscr{S} \subset \mathrm{PU}(d)$, and any $t \geqslant 1$,*

$$\sum_{x,y \in \mathscr{S}} w(x) w(y) |\mathrm{tr}[U(x)^{\dagger} U(y)]|^{2t} \geqslant \gamma(t, d), \tag{5.5}$$

with equality if and only if $(\mathscr{S}, w)$ is a weighted t-design.

Proof. Defining $S := \sum_{x \in \mathscr{S}} w(x) U(x)^{\otimes t} \otimes (U(x)^{\otimes t})^{\dagger} - \int_{\mathrm{PU}(d)} \mathrm{d}\mu(x) U(x)^{\otimes t} \otimes (U(x)^{\otimes t})^{\dagger}$ we see that

$$\begin{aligned} 0 \leqslant \mathrm{tr}(S^{\dagger} S) = &\sum_{x,y \in \mathscr{S}} w(x) w(y) |\mathrm{tr}[U(x)^{\dagger} U(y)]|^{2t} - 2 \sum_{x \in \mathscr{S}} w(x) \int_{\mathrm{PU}(d)} \mathrm{d}\mu(y) |\mathrm{tr}[U(x)^{\dagger} U(y)]|^{2t} \\ &+ \int_{\mathrm{PU}(d)} \mathrm{d}\mu(x) \int_{\mathrm{PU}(d)} \mathrm{d}\mu(y) |\mathrm{tr}[U(x)^{\dagger} U(y)]|^{2t} \qquad (5.6) \\ = &\sum_{x,y \in \mathscr{S}} w(x) w(y) |\mathrm{tr}[U(x)^{\dagger} U(y)]|^{2t} - \int_{\mathrm{PU}(d)} \mathrm{d}\mu(x) |\mathrm{tr}[U(x)]|^{2t} \qquad (5.7) \end{aligned}$$

with equality if and only if $S = 0$, which is the defining property of a t-design. □

This theorem allows us to check whether a weighted subset of $\mathrm{PU}(d)$ forms a t-design by considering only the 'angles' between the supposed design elements. It also shows that t-designs can be found numerically by parametrizing a weighted set and minimizing the LHS of equation (5.5). The lower bound can be considered a variation on the Welch bound [26].

The constant γ was calculated by Diaconis and Shahshahani [27] for $d \geqslant t$, in which case $\gamma(t, d) = t!$, and by Rains [28] in general. It is the number of permutations $\sigma \in \mathrm{S}_t$ (the symmetric group [29]) such that $(\sigma(1), \sigma(2), \ldots, \sigma(t))$ has no increasing subsequence of length greater than d. Thus, for example, $\gamma(1, d) = 1$ and $\gamma(2, d) = 2$ for all $d \geqslant 2$.

A unitary 1-design must satisfy

$$\sum_{x \in \mathscr{D}} w(x) U(x) \otimes U(x)^\dagger = \int_{\mathrm{PU}(d)} \mathrm{d}\mu(x) U(x) \otimes U(x)^\dagger = \frac{1}{d} T, \tag{5.8}$$

where the RHS of equation (5.1) is now explicitly evaluated (simply consider a matrix component of the integral in the standard product basis and use Schur's lemma). Since T has eigenvalues of 1 and -1, respectively, on the symmetric and antisymmetric subspaces of $\mathbb{C}^d \otimes \mathbb{C}^d$, and thus $\mathrm{rank}(T) = d^2$, we must have $|\mathscr{D}| \geqslant d^2$ with equality only if $\mathscr{D}$ is an (orthogonal) unitary operator basis, i.e. $\mathrm{tr}[U(x)^\dagger U(y)] = 0$ for all $x \neq y \in \mathscr{D}$, and $w(x) = 1/|\mathscr{D}|$. This fact is more apparent when equation (5.8) is rewritten in terms of superoperators

$$\sum_{x \in \mathscr{D}} w(x) |U(x))(U(x)| = \frac{1}{d} \mathbf{I}. \tag{5.9}$$

In this form it is clear that unitary 1-designs are equivalent to tight unitary frames [5, 30]. The unitary operators with matrix elements [30]

$$\langle j | U_m | k \rangle := \frac{1}{\sqrt{d}} \exp\left[\frac{2\pi \mathrm{i} jk}{d} + \frac{2\pi \mathrm{i}(j + kd)m}{n} \right], \tag{5.10}$$

for $m = 0, \ldots, n-1$, provide explicit examples of (unweighted) unitary 1-designs for all $n = |\mathscr{D}| \geqslant d^2$.

To treat the general case, for an arbitrary permutation $\sigma \in \mathrm{S}_n$, define the *permutation operator*

$$P(\sigma) = P_{\sigma(1)\sigma(2)\cdots\sigma(n)} := \sum_{j_1, j_2, \ldots, j_n} |j_1\rangle\langle j_{\sigma(1)}| \otimes |j_2\rangle\langle j_{\sigma(2)}| \otimes \cdots \otimes |j_n\rangle\langle j_{\sigma(n)}|, \tag{5.11}$$

which acts on $(\mathbb{C}^d)^{\otimes n}$ by permuting its subsystems accordingly,

$$P_{k_1 k_2 \ldots k_n} |\psi_{k_1}\rangle \otimes |\psi_{k_2}\rangle \otimes \cdots \otimes |\psi_{k_n}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle, \tag{5.12}$$

and in terms of operators,

$$P_{k_1 k_2 \ldots k_n} \left(A_{k_1} \otimes A_{k_2} \otimes \cdots \otimes A_{k_n} \right) P_{k_1 k_2 \ldots k_n}{}^\dagger = A_1 \otimes A_2 \otimes \cdots \otimes A_n. \tag{5.13}$$

The composition of permutation operators then follows that for permutations, $P(\sigma)P(\tau) = P(\sigma\tau)$, which means $P(\sigma)^\dagger = P(\sigma^{-1})$. Note that $P_{21} = T$.

In general, the RHS of equation (5.1) can be integrated explicitly using group theoretical methods [31, 32]. The result is

$$\int_{\mathrm{PU}(d)} \mathrm{d}\mu(x) U(x)^{\otimes t} \otimes (U(x)^{\otimes t})^\dagger = \sum_{\sigma, \tau \in \mathrm{S}_t} \mathrm{Wg}(d, t, \sigma\tau^{-1}) P_{\tau(1)+t, \ldots, \tau(t)+t, \sigma^{-1}(1), \ldots, \sigma^{-1}(t)}, \tag{5.14}$$

where

$$\mathrm{Wg}(d, t, \sigma) := \frac{1}{t!^2} \sum_{\substack{\lambda \vdash t \\ l(\lambda) \leqslant d}} \frac{\chi^\lambda(1)^2 \chi^\lambda(\sigma)}{s_{\lambda, d}(1)} \tag{5.15}$$

is called the *Weingarten function*. Here the sum is over all partitions $\lambda = (\lambda_1, \ldots, \lambda_t)$ of the integer t (i.e. nonincreasing sequences of non-negative integers summing to t) with

length $l(\lambda) \leqslant d$, where $l(\lambda) := \max_{\lambda_j>0} j$. The character on the conjugacy class K_λ of S_t corresponding to $\lambda \vdash t$ is denoted by χ^λ and we take $s_{\lambda,d}(1) = s_{\lambda,d}(1,\ldots,1)$ for the Schur function $s_{\lambda,d}(x_1,\ldots,x_d)$ [29]. For any partition $\lambda \vdash t$ one has

$$s_{\lambda,d}(1) = \frac{1}{t!}\sum_{\mu\vdash t} d^{l(\mu)}\chi^\lambda(\mu)|\mathrm{K}_\mu|, \tag{5.16}$$

and in particular, $s_{(1,1),d}(1) = d(d+1)/2$ and $s_{(2,0),d}(1) = d(d-1)/2$. This means $\mathrm{Wg}(d,2,(1,1)) = 1/(d^2-1)$ and $\mathrm{Wg}(d,2,(2,0)) = -1/d(d^2-1)$, giving

$$\int_{\mathrm{PU}(d)} \mathrm{d}\mu(x) U(x)\otimes U(x)\otimes U(x)^\dagger \otimes U(x)^\dagger$$
$$= \frac{1}{d^2-1}\left(P_{3412}+P_{4321}\right) - \frac{1}{d(d^2-1)}\left(P_{4312}+P_{3421}\right). \tag{5.17}$$

Thus our definition of a unitary 2-design [equation (5.1)] can be rewritten as

$$\sum_{x\in\mathscr{D}} w(x) U(x)\otimes U(x)\otimes U(x)^\dagger \otimes U(x)^\dagger$$
$$= \frac{1}{d^2-1}\left(P_{3412}+P_{4321}\right) - \frac{1}{d(d^2-1)}\left(P_{4312}+P_{3421}\right). \tag{5.18}$$

The following is partly due to Gross *et al* [11, theorem 2].

Theorem 5.5. *Let* $(\mathscr{D}, w)$, $\mathscr{D} \subset \mathrm{PU}(d)$, *be a weighted 2-design. Then*

$$|\mathscr{D}| \geqslant (d^2-1)^2 + 1, \tag{5.19}$$

with equality only if $w(x) = 1/|\mathscr{D}|$ *and*

$$|\mathrm{tr}[U(x)^\dagger U(y)]|^2 = 1 - \frac{1}{d^2-1}, \tag{5.20}$$

for all $x \neq y \in \mathscr{D}$.

Proof. Multiplying equation (5.18) on the left by P_{2341} and on the right by ${P_{2341}}^\dagger = P_{4123}$ implies

$$\sum_{x\in\mathscr{D}} w(x) U(x)^\dagger\otimes U(x)\otimes U(x)\otimes U(x)^\dagger$$
$$= \frac{1}{d^2-1}\left(P_{3412}+P_{2143}\right) - \frac{1}{d(d^2-1)}\left(P_{3142}+P_{2413}\right), \tag{5.21}$$

given equation (5.13) and since, for example, $P_{2341}P_{3412}P_{4123} = P_{2341}P_{2341} = P_{3412}$. Now multiply this equation on the right by $A\otimes I\otimes I$, where $A \in \mathrm{End}(\mathbb{C}^d)\otimes\mathrm{End}(\mathbb{C}^d)$, and trace out the first pair of subsystems. The result can be written in terms of a superoperator

$$\mathcal{S}(A) := \sum_{x\in\mathscr{D}} w(x)\,\mathrm{tr}[(U(x)^\dagger\otimes U(x))A]U(x)\otimes U(x)^\dagger \tag{5.22}$$
$$= \frac{A+\mathrm{tr}(AT)T}{d^2-1} - \frac{(I\otimes\mathrm{tr}_1(AT))T + (\mathrm{tr}_2(AT)\otimes I)T}{d(d^2-1)}, \tag{5.23}$$

by rewriting the permutation operators explicitly in terms of their definition [equation (5.11)] and simplifying.

Now let $\{E_k\}_{k=0}^{d^2-1}$ be an orthonormal operator basis for $\mathrm{End}(\mathbb{C}^d)$ with the choice $E_0 = I/\sqrt{d}$. The remaining operators $E_1,\ldots,E_{d^2-1}$ then span the (d^2-1)-dimensional

subspace of traceless operators $\mathrm{tr}(E_k) = \sqrt{d}\,\mathrm{tr}(E_0^\dagger E_k) = 0$ for all $k > 0$. Consider the action of $\mathcal{S}$ on $(E_j \otimes E_k)T$:

$$\mathcal{S}((E_j \otimes E_k)T) = \frac{1}{d^2-1}((E_j \otimes E_k)T + d^2\delta_{j0}\delta_{k0}(E_0 \otimes E_0)T - \delta_{j0}(E_0 \otimes E_k)T - \delta_{0k}(E_j \otimes E_0)T) \tag{5.24}$$

$$= \begin{cases} (E_0 \otimes E_0)T, & j = k = 0; \\ (E_j \otimes E_k)T/(d^2-1), & j,k > 0; \\ 0, & \text{otherwise}, \end{cases} \tag{5.25}$$

identifying $I = \sqrt{d}E_0$. Thus the d^4 orthonormal operators $(E_j \otimes E_k)T$ diagonalize $\mathcal{S}$ and, in particular, $\mathrm{rank}'(\mathcal{S}) = (d^2-1)^2 + 1$ (ordinary rank). But we must have $|\mathscr{D}| \geqslant \mathrm{rank}'(\mathcal{S})$, which is equation (5.19).

If $|\mathscr{D}| = \mathrm{rank}'(\mathcal{S})$ then $\{U(x) \otimes U(x)^\dagger\}_{x\in\mathscr{D}}$ is necessarily a linearly independent set. Fixing $y \in \mathscr{D}$ and considering $\mathcal{S}(U(y) \otimes U(y)^\dagger)$ shows that

$$(d^2-1)\sum_{x\in\mathscr{D}} w(x)|\mathrm{tr}[U(x)^\dagger U(y)]|^2 U(x) \otimes U(x)^\dagger = U(y) \otimes U(y)^\dagger + \left(d - \frac{2}{d}\right)T, \tag{5.26}$$

which, upon setting $T = d\sum_{x\in\mathscr{D}} w(x)U(x) \otimes U(x)^\dagger$ (equation (5.8)), can be rewritten as

$$\{((d^2-1)^2+1)w(y) - 1\}U(y) \otimes U(y)^\dagger + \sum_{x\neq y} w(x)\{(d^2-1)|\mathrm{tr}[U(x)^\dagger U(y)]|^2 - d^2 + 2\}U(x) \otimes U(x)^\dagger = 0. \tag{5.27}$$

When $|\mathscr{D}| = (d^2-1)^2+1$ linear independence thus requires $(d^2-1)|\mathrm{tr}[U(x)^\dagger U(y)]|^2 = d^2-2$ for all $x \neq y$ and $((d^2-1)^2+1)w(y) = |\mathscr{D}|w(y) = 1$. The same is true for all $y \in \mathscr{D}$. □

In general, for each positive integer t and d, we would like to know the quantity $N(t,d)$, which we use to denote the minimum number of unitaries needed to construct a weighted t-design in $\mathrm{PU}(d)$, or less ambitiously, bounds on this quantity. This is a difficult problem. A general lower bound, however, might be obtainable from the theory of Levenshtein [33, 34].

Our own numerical searches have not revealed the existence of 2-designs achieving $|\mathscr{D}| = (d^2-1)^2+1$ and Gross *et al* [11] have conjectured their nonexistence. If such designs did exist then they might be dubbed *tight* designs, which is standard terminology in the theory of t-designs [7] (but unrelated to the concept of tight frames). As noted in theorem 5.5, tight $\mathrm{PU}(d)$ 2-designs are necessarily equiangular. They are analogous to tight $\mathbb{C}P^{d-1}$ 2-designs, i.e. symmetric informationally complete POVMs (SIC-POVMs) [35], which in contrast are conjectured to exist in all dimensions. The analogy to a complete family of mutually unbiased bases (MUBs) [36, 37], however, does exist in certain dimensions.

A subset $\{U_j\}_{j=0}^{d^2-1} \subset \mathrm{U}(d)$ is a unitary operator basis for $\mathrm{End}(\mathbb{C}^d)$ if $\mathrm{tr}\big(U_j{}^\dagger U_k\big) = d\delta_{jk}$ for all $0 \leqslant j,k \leqslant d^2-1$. In analogy with the case of vector bases, we call a pair of unitary operator bases, $\{U_j\}_{j=0}^{d^2-1}$ and $\{V_k\}_{k=0}^{d^2-1}$, *mutually unbiased* if

$$\left|\mathrm{tr}\big(U_j{}^\dagger V_k\big)\right|^2 = 1 \tag{5.28}$$

for all $0 \leqslant j,k \leqslant d^2-1$. Define the embedding $\vartheta : \mathrm{U}(d) \hookrightarrow \mathrm{H}_0^{\mathrm{uc}} \cong \mathbb{R}^{(d^2-1)^2}$ by

$$|\vartheta(U)) := \mathbf{\Pi}_0||U\rangle\langle U|) = ||U\rangle\langle U| - I/d^2), \tag{5.29}$$

where $|U\rangle\langle U|$, $\mathbf{\Pi}_0$ and $\mathrm{H}_0^{\mathrm{uc}}$ are defined in equations (2.2), (2.14) and (2.17), respectively. The set $\{\vartheta(U_j)\}_{j=0}^{d^2-1}$ then specifies the vertices of a regular simplex in the (d^2-1)-dimensional

subspace of $\mathrm{H}_0^{\mathrm{uc}}$ for which its members span. Mutually unbiased bases correspond to orthogonal subspaces,

$$(\vartheta(U_j)|\vartheta(V_k)) = |\langle U_j|V_k\rangle|^2 - \frac{1}{d^2} = \frac{1}{d^2}\left|\mathrm{tr}\left(U_j{}^{\dagger}V_k\right)\right|^2 - \frac{1}{d^2} = 0, \tag{5.30}$$

of which, there can be at most $\left(\dim \mathrm{H}_0^{\mathrm{uc}}\right)/(d^2-1) = d^2-1$ many. A set of d^2-1 unitary operator bases with the property that each pair is mutually unbiased is thus called a *complete* set of mutually unbiased unitary-operator bases (MUUBs).

Now consider an arbitrary family of subsets, $\mathscr{B}_0, \ldots, \mathscr{B}_{m-1} \subset \mathrm{PU}(d)$, where each member $\mathscr{B}_a = \left\{e_j^a\right\}_{j=0}^{d^2-1}$ specifies a unitary operator basis $\left\{U\left(e_j^a\right)\right\}_{j=0}^{d^2-1}$ and is appointed a positive weight w_a. By theorem 5.4, if

$$\sum_{a,b=0}^{m-1} w_a w_b \sum_{j,k=0}^{d^2-1} \left|\mathrm{tr}\left[U\left(e_j^a\right)^{\dagger} U\left(e_k^b\right)\right]\right|^4 = 2, \tag{5.31}$$

then their union $\mathscr{D} = \cup_a \mathscr{B}_a$ forms a weighted 2-design with weight $w(x) = \sum_a w_a 1_{\mathscr{B}_a}(x)$. In the context of quantum process tomography it is desirable for the weight to remain constant across elements of the same basis (so the POVM that the design specifies can be implemented by a series of orthogonal measurements). We have thus made this a requirement. The set indicator function, $1_{\mathscr{S}}(x) := 1$ if $x \in \mathscr{S}$ and 0 otherwise, is used to take care of any multiplicity across different bases. Note that the normalization of $w(x)$ implies normalization of the basis weights: $\sum_a w_a = 1/d^2$.

It is straightforward to confirm (via equation (5.31)) that a complete set of MUUBs forms a unitary 2-design when $w_a = 1/md^2$. The following theorem shows that such sets are optimal, in that we always need $m \geqslant d^2-1$ unitary operator bases to construct a weighted 2-design, with equality only if the bases are mutually unbiased.

Theorem 5.6. *Let $d > 1$ and let $\mathscr{B}_0, \ldots, \mathscr{B}_{m-1} \subset \mathrm{PU}(d)$ specify a family of unitary operator bases for $\mathrm{End}(\mathbb{C}^d)$, where the union $\mathscr{D} = \cup_a \mathscr{B}_a$ forms a weighted 2-design with weight function $w(x) = \sum_a w_a 1_{\mathscr{B}_a}(x)$ for some choice of basis weights $w_0, \ldots, w_{m-1} > 0$. Then $m \geqslant d^2-1$ with equality only if $w_a = 1/md^2$ for all a and the bases are pairwise mutually unbiased.*

Proof. Theorem 5.5 with $|\mathscr{D}| = md^2$ immediately shows that we require $m \geqslant d^2 - 2 + 2/d^2$, which means $m \geqslant d^2-1$ whenever $d > 1$. In the case of equality, note that by theorem 5.4 (or equation (5.31)) we require

$$d^6 \sum_a {w_a}^2 + \sum_{a\neq b} w_a w_b \sum_{j,k} \left(\lambda_{jk}^{ab}\right)^2 = 2, \tag{5.32}$$

where we have defined the positive numbers $\lambda_{jk}^{ab} := \left|\mathrm{tr}\left[U\left(e_j^a\right)^{\dagger} U\left(e_k^b\right)\right]\right|^2$. Moreover, theorem 5.4 implies that the LHS of equation (5.32) is minimal with respect to the variables w_a and λ_{jk}^{ab} under the appropriate constraints, two of which are $\sum_a w_a = 1/d^2$ and

$$\sum_{j,k} \lambda_{jk}^{ab} = \sum_{j,k} \left|(U\left(e_j^a\right)|U\left(e_k^b\right))\right|^2 = d^2\,\mathrm{Tr}(\mathbf{I}\cdot\mathbf{I}) = d^4, \tag{5.33}$$

since all unitary operator bases satisfy $\sum_j |U(e_j))(U(e_j)| = d\mathbf{I}$. We will now minimize the LHS of equation (5.32) under these two constraints. The minimum of $\sum_{j,k}\left(\lambda_{jk}^{ab}\right)^2$ subject to equation (5.33) occurs only when $\lambda_{jk}^{ab} = 1$ for all $0 \leqslant j,k \leqslant d^2-1$, i.e., when $\mathscr{B}_a$ and $\mathscr{B}_b$ are mutually unbiased. Then the LHS of equation (5.32) reduces to

$$d^6 \sum_a w_a^2 + d^4 \sum_{a\neq b} w_a w_b = d^4(d^2-1)\sum_a w_a^2 + 1, \tag{5.34}$$

and here the minimum (under $\sum_a w_a = 1/d^2$) occurs only when $w_a = 1/md^2$ for all $0 \leqslant a \leqslant m-1$. With this value, equation (5.34) reduces to the RHS of equation (5.32) when $m = d^2 - 1$. Equality in equation (5.32) thus requires the bases to be pairwise mutually unbiased and $w_a = 1/md$ whenever $m = d^2 - 1$. □

Theorem 5.6 is the equivalent of ([38, theorem 3.2] for the case of unitary designs. Many examples of unweighted unitary 2-designs were described by Gross *et al* [11]. Of these, the *Clifford designs* were found closest to optimal. These are sets of unitary operator bases which form subgroups of the projective Clifford group PC(d) [39, 40] and have cardinalities $|\mathscr{D}| = kd^2(d^2-1)$ for some integer k. When $k = 1$, Clifford designs are known to exist in dimensions $d = 2, 3, 5, 7, 11$ [41]. By theorem 5.6, each of these examples must be the union of a complete set of MUUBs. Although no unweighted unitary 2-designs of smaller size were found by Gross *et al* [11], weighted unitary 2-designs can surpass this record. This is the case for PU(2) 2-designs, which are described in detail next.

5.1. PU(2) t-designs

In dimension 2, unitary designs are equivalent to real projective designs, which in turn are equivalent to antipodal spherical designs. To see this, simply note that $\mathrm{PU}(2) \cong \mathbb{R}P^3$ through the relation

$$\mathrm{e}^{\mathrm{i}\phi} U = r_0 I + \mathrm{i}(r_1 X + r_2 Y + r_3 Z), \tag{5.35}$$

where $X := \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $Y := \left(\begin{smallmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{smallmatrix}\right)$ and $Z := \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ are the Pauli matrices. Each unit vector $(r_0, r_1, r_2, r_3) \in \mathbb{R}^4$ specifies a line in $\mathbb{R}P^3$, and through equation (5.35), an equivalence class of unitaries $U \in \mathrm{U}(2)$ differing only be a phase factor. Under this map each t-design in PU(2) gives a t-design in $\mathbb{R}P^3$ and vice versa. This is because distances are preserved: $|\mathrm{tr}(U^\dagger V)|^2 = 4\langle r|s\rangle^2$ where $\langle r|s\rangle := \sum_k r_k s_k$ and r (respectively, s) corresponds to U (respectively, V) through equation (5.35). Theorem 5.4 then transforms to the equivalent for real projective designs. This relationship also gives

$$\gamma(t, 2) = \frac{(2t)!}{t!(t+1)!}. \tag{5.36}$$

Real projective designs are rarely studied in the literature. It is well known, however, that t-designs in $\mathbb{R}P^{n-1}$ are equivalent to antipodal $(2t+1)$-designs in S^{n-1} with twice as many points (assuming antipodal pairs are appointed the same weight). The antipodal points of the spherical design are simply the intersections between the lines of the real projective design and the unit sphere. Additionally, an antipodal spherical $(2t+1)$-design can be created from $(2t)$-design by simply appending the antipodal points to the design: if $\mathscr{D}$ is a $(2t)$-design in S^{n-1} then $\mathscr{D} \cup (-\mathscr{D})$ is an antipodal $(2t+1)$-design in S^{n-1}.

The above relationships can be used to translate known results in the literature to the case of unitary designs. For example, the lower bound of Delsarte *et al* [7] on the number of points needed to construct a $(2t+1)$-design in S^3 shows that

$$|\mathscr{D}| \geqslant \tfrac{1}{6}(t+1)(t+2)(t+3) \tag{5.37}$$

for a t-design in PU(2), with equality only if the design is unweighted [34], i.e. $w(x) = 1/|\mathscr{D}|$. A design which achieves this bound is generally called *tight*. It is known, however, that tight S^3 $(2t+1)$-designs exist only for the trivial $t = 1$ case [7, 42, 43] (see [44] for a summary). Thus we can increase the RHS of equation (5.37) by 1 when $t > 1$. Further bounds on the cardinality of a PU(2) t-design are summarized in table 1.

Table 1. Known bounds on $N(t, 2)$, the minimum cardinality of a weighted t-design in PU(2). The Delsarte lower bound (equation (5.37)) is included as a reference point, and for completeness, bounds on the minimum cardinality for an unweighted design are included in parentheses.

t	Delsarte	$N(t,2) \geqslant$	$N(t,2) \leqslant$
2	10	11[a] (12[b])	11[c] (12[d])
3	20	21[a]	23[e] (24[f])
4	35	37[g]	43[e]
5	56	60[g]	60[h]
6	84	85[a] (89[i])	
7	120	134[g]	264[j]
8	165	166[a] (180[i])	
9	220	250[g]	360[k]
10	286	287[a] (318[i])	

[a] No tight $S^3(2t+1)$-designs exist for $t > 1$ [7, 42, 43].
[b] No antipodal unweighted 22-point S^3 5-designs exist [45].
[c] A weighted 11-point PU(2) 2-design exists [equation (5.38)].
[d] The 24 vertices of the 24-cell form an antipodal unweighted S^3 5-design [46], and thus also an unweighted PU(2) 2-design. This design is a minimal subgroup of PC(2).
[e] A weighted 23-point S^3 6-design and a weighted 43-point S^3 8-design exist [47].
[f] The projective Clifford group PC(2) is an unweighted PU(2) 3-design. The corresponding S^3 7-design is formed by the vertices of 2 copies of the 24-cell [46].
[g] The linear programming bounds for weighted $S^3(2t+1)$-designs of [48].
[h] The 120 vertices of the 600-cell form an antipodal unweighted S^3 11-design. This is the unique minimal unweighted S^3 11-design [49, 50].
[i] Yudin's bound [51] on unweighted spherical designs gives $|\mathscr{D}| \geqslant \pi/(\pi - 2x\sqrt{1-x^2} - 2\arcsin x)$ for unweighted PU(2) t-designs, where x is the largest zero of the Jacobi polynomial $P_{2t+1}^{(3/2,3/2)}(x)$.
[j] An antipodal weighted 528-point S^3 15-design can be constructed from shells of a Euclidean lattice [48].
[k] The union of the 120 vertices and the 600 face centres of the 600-cell form a weighted antipodal S^3 19-design [52]. The vertices have weight $1/504$ and the faces have weight $2/1575$.

The first row of table 1 corresponds to $t = 2$, which we now explain in detail. The Delsarte bound (equation (5.37)) shows that PU(2) 2-designs must have at least ten points. However we can increase this bound to 11 since there are no tight PU(2) 2-designs. The 11 columns of the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & a & a & a & a & a & a \\ 0 & a & -a & a & a & b & -b & 0 & 0 & 0 & 0 \\ 0 & a & a & -a & a & 0 & 0 & b & -b & 0 & 0 \\ 0 & a & a & a & -a & 0 & 0 & 0 & 0 & b & -b \end{bmatrix}, \tag{5.38}$$

where $a = 1/\sqrt{3}$ and $b = \sqrt{2/3}$, specify a weighted PU(2) 2-design (through equation (5.35)) which achieves the bound. The weight appointed to the first column is $1/16$ while the remaining all have weight $3/32$. Reznick [45] has shown that there are no antipodal unweighted S^3 5-designs with 22 points. Thus this design is necessarily weighted. The 24 vertices of the 24-cell form an antipodal unweighted S^3 5-design [46], and thus also an unweighted PU(2) 2-design with 12 points. The elements of this design correspond to the subgroup $\langle HR, R^2 \rangle$ of the Clifford group $\mathrm{C}(2) = \langle H, R \rangle$, where $H = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$ and $R = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{smallmatrix}\right)$. It is formed by the union of a complete set of MUUBs and is the first in the family of Clifford designs.

Finally, Shamsiev's explicit constructions of antipodal weighted spherical designs [53, theorem 1] show that weighted PU(2) t-designs with $(t+1)^3/2$ points exist for all odd t. This

upper bound on the cardinality together with the Delsarte lower bound (equation (5.37)) means that $N(t, 2) = \Theta(t^3)$.

6. Optimal ancilla-assisted quantum process tomography

We now return to our immediate task of optimizing the measurements used for ancilla-assisted quantum process tomography. Throughout this section we assume that either $\mathscr{Q} = \mathrm{Q}(\mathbb{C}^d \otimes \mathbb{C}^d)$, $\mathscr{Q}^{\mathrm{gc}}$ or $\mathscr{Q}^{\mathrm{uc}}$. The results of section 4 are first summarized for these specific cases.

Recall that our error for tomographic reconstructions of quantum states was defined in terms of the Hilbert–Schmidt distance (equation (4.4)),

$$e^{(F,Q)}(\rho) := \mathop{\mathrm{E}}_{y_1,\ldots,y_N} [\|\rho - \hat{\rho}(y_1, \ldots, y_N)\|^2], \tag{6.1}$$

where $\hat{\rho}$ is the linear tomographic estimate of ρ given N measurement outcomes $y_1, \ldots, y_N$ (equations (4.2) and (4.3)). When applied to the output states of quantum channels, with fixed input $\rho_{\mathrm{i}} = |I\rangle\langle I|$, this distance measure naturally induces the analogous Hilbert–Schmidt distance for superoperators (see the appendix),

$$\|\mathcal{E} - \hat{\mathcal{E}}\| = \|\rho - \hat{\rho}\|, \tag{6.2}$$

where $\rho = (\mathcal{E} \otimes \mathcal{I})(\rho_{\mathrm{i}})$ and $\hat{\rho} = (\hat{\mathcal{E}} \otimes \mathcal{I})(\rho_{\mathrm{i}})$ through the Jamiołkowski isomorphism (equation (2.3)), and $\|\mathcal{S}\| := \sqrt{\mathrm{Tr}(\mathcal{S}^\dagger \mathcal{S})}$ for any superoperator $\mathcal{S}$. Although there are more appropriate distance measures for quantum channels, which properly reflect the probabilistic interpretation of a quantum state, the Hilbert–Schmidt distance is the most natural choice for linear tomographic reconstructions of quantum states.

Now setting $\rho = \rho(\sigma, U) := U\sigma U^\dagger$ for some fixed output state $\sigma \in \mathscr{Q}$, recall that we defined the average (equation (4.16)) and worst-case (equation (4.32)) error over different Hilbert-space orientations $U_{\mathrm{s}}, U_{\mathrm{a}} \in \mathrm{U}(d)$,

$$e^{(F,Q)}_{\mathrm{av}}(\sigma) := \int\!\!\int_{\mathrm{U}(d)} \mathrm{d}\mu(U_{\mathrm{s}})\mathrm{d}\mu(U_{\mathrm{a}})\, e^{(F,Q)}(\rho(\sigma, U_{\mathrm{s}} \otimes U_{\mathrm{a}})); \tag{6.3}$$

$$e^{(F,Q)}_{\mathrm{wc}}(\sigma) := \sup_{U_{\mathrm{s}},U_{\mathrm{a}} \in \mathrm{U}(d)} e^{(F,Q)}(\rho(\sigma, U_{\mathrm{s}} \otimes U_{\mathrm{a}})). \tag{6.4}$$

Finally, recalling that the subsets $\mathscr{Q}^{\mathrm{gc}}$ and $\mathscr{Q}^{\mathrm{uc}}$ respectively span $\delta' = d^2(d^2 - 1) + 1$ and $\delta' = (d^2 - 1)^2 + 1$ dimensions of $\mathrm{H}(\mathbb{C}^d \otimes \mathbb{C}^d)$, the following corollary restates theorem 4.4 and corollary 4.5 for these special cases of interest.

Corollary 6.1. *Let $F : \mathfrak{B}(\mathscr{X}) \to \mathrm{H}(\mathbb{C}^d \otimes \mathbb{C}^d)$ be a POVM which is informationally complete w.r.t. $\mathscr{Q} \subseteq \mathrm{Q}(\mathbb{C}^d \otimes \mathbb{C}^d)$. Then for any fixed quantum state $\sigma \in \mathscr{Q}$,*

$$e^{(F,Q)}_{\mathrm{wc}}(\sigma) \geqslant e^{(F,Q)}_{\mathrm{av}}(\sigma) \geqslant \begin{cases} \dfrac{1}{N}(d^4 + d^2 - 1 - \mathrm{tr}(\sigma^2)), & \text{if} \quad \mathscr{Q} = \mathrm{Q}(\mathbb{C}^d \otimes \mathbb{C}^d); \\ \dfrac{1}{N}(d^4 - d^2 + 1/d^2 - \mathrm{tr}(\sigma^2)), & \text{if} \quad \mathscr{Q} = \mathscr{Q}^{\mathrm{gc}}; \\ \dfrac{1}{N}(d^4 - 3d^2 + 3 - \mathrm{tr}(\sigma^2)), & \text{if} \quad \mathscr{Q} = \mathscr{Q}^{\mathrm{uc}}, \end{cases} \tag{6.5}$$

for all reconstruction OVDs Q. Furthermore, equality in the RHS of equation (6.5) occurs if and only if $Q = R$ and F is a tight rank-1 POVM with $\mathrm{supp}(\mathcal{F}) = \mathrm{span}(\mathscr{Q})$*, in which case we also have equality in the LHS of equation (6.5).*

Tight rank-1 POVMs thus describe the class of optimal measurements for *linear* ancilla-assisted quantum process tomography in both an average and worst-case sense. But do such measurements exist? When $\mathscr{Q} = \mathrm{Q}(\mathbb{C}^d \otimes \mathbb{C}^d)$, which is the class of output states of general quantum operations (and included in corollary 6.1 for comparison), we recover the results of [5]. Here it was found that tight rank-1 POVMs exist and are in fact equivalent to weighted complex projective 2-designs.

Now consider the case $\mathscr{Q} = \mathscr{Q}^{\mathrm{uc}}$. Our condition for a tight rank-1 POVM (equation (4.28)) with $D = d^2$, $\delta = (d^2-1)^2 + 1$ and

$$\mathbf{\Pi}_{\mathcal{F}} = \mathbf{\Pi}_{\mathscr{Q}^{\mathrm{uc}}} = |\lambda_0 \otimes \lambda_0)(\lambda_0 \otimes \lambda_0| + \sum_{j,k>0} |\lambda_j \otimes \lambda_k)(\lambda_j \otimes \lambda_k|, \tag{6.6}$$

becomes

$$\mathcal{F} = \int_{\mathscr{X}} \mathrm{d}\tau(x) |P(x))(P(x)| = |\lambda_0 \otimes \lambda_0)(\lambda_0 \otimes \lambda_0| + \frac{1}{d^2-1} \sum_{j,k>0} |\lambda_j \otimes \lambda_k)(\lambda_j \otimes \lambda_k|, \tag{6.7}$$

using the orthonormal Hermitian operator basis $\{\lambda_k\}_{k=0}^{d^2-1}$ of section 2 (see above equation (2.9)). Equivalently, under the isomorphism $|A)(B| \leftrightarrow A \otimes B^\dagger$ we can rewrite this last form as

$$\int_{\mathscr{X}} \mathrm{d}\tau(x) P(x) \otimes P(x) = (\lambda_0 \otimes \lambda_0) \otimes (\lambda_0 \otimes \lambda_0) + \frac{1}{d^2-1} \sum_{j,k>0} (\lambda_j \otimes \lambda_k) \otimes (\lambda_j \otimes \lambda_k). \tag{6.8}$$

Now multiplying on the left by P_{1432}, taking the trace, and applying the easily confirmed identity $\mathrm{tr}[P_{1432}(A \otimes B \otimes C \otimes D)] = \mathrm{tr}(A)\,\mathrm{tr}(C)\,\mathrm{tr}(BD)$, we find that the types of tight rank-1 POVMs corresponding to $\mathscr{Q}^{\mathrm{uc}}$ must satisfy

$$\int_{\mathscr{X}} \mathrm{d}\tau(x)\, \mathrm{tr}_{\mathrm{s}}[\{\mathrm{tr}_{\mathrm{a}}[P(x)]\}^2] = d. \tag{6.9}$$

We know that $\mathrm{tr}_{\mathrm{s}}[\{\mathrm{tr}_{\mathrm{a}}[P]\}^2] \leqslant 1/d$, however, with equality only if $P = |U\rangle\langle U|$ (via equation (2.2)), a maximally entangled state. Thus, since the normalization $\int_{\mathscr{X}} \mathrm{d}\tau(x) = D = d^2$ must be adhered to, equation (6.9) can be satisfied only if $P(x) = |U(x)\rangle\langle U(x)|$, τ-almost everywhere, for some function $U : \mathscr{X} \to \mathrm{U}(d)$.

We have established that all tight rank-1 POVMs corresponding to $\mathscr{Q}^{\mathrm{uc}}$ have POVDs in the form $P(x) = |U(x)\rangle\langle U(x)|$ where $U : \mathscr{X} \to \mathrm{U}(d)$. It is thus natural to take $\mathscr{X} \subseteq \mathrm{PU}(d)$ and let $U(x)$ denote a representative from the equivalence class of unitaries $x \in \mathrm{PU}(d)$ (as in section 5). We will henceforth assume that this is the case. Now note that $\mathrm{Tr}(\mathcal{F}^2) = 2$ under equation (6.7). This means

$$\frac{1}{d^4} \int\!\!\int_{\mathscr{X}} \mathrm{d}\tau(x)\, \mathrm{d}\tau(y) |\mathrm{tr}[U(x)^\dagger U(y)]|^4 = 2, \tag{6.10}$$

given that $|(P(x)|P(y))|^2 = |\langle U(x)|U(y)\rangle|^4 = |\mathrm{tr}[U(x)^\dagger U(y)]|^4/d^4$. In particular, if $\mathscr{X}$ is a finite set, then by theorem 5.4, $\mathscr{X}$ must be a weighted unitary 2-design with weight function $w(x) = \tau(x)/d^2$. In fact, theorem 5.4 could easily be extended to any subset $\mathscr{X} \subseteq \mathrm{PU}(d)$ with the condition for equality in equation (5.5) (when $t = 2$) replaced by equation (6.11) in the following proposition (see, e.g., [5, theorem 6]).

Proposition 6.2. *Let* $F : \mathfrak{B}(\mathscr{X}) \to \mathrm{H}(\mathbb{C}^d \otimes \mathbb{C}^d)$ *be a POVM with* $\mathrm{supp}(\mathcal{F}) = \mathrm{span}(\mathscr{Q}^{\mathrm{uc}})$ *and assume* $\mathscr{X} \subseteq \mathrm{PU}(d)$. *Then F is a tight rank-1 POVM if and only if* $P(x) = |U(x)\rangle\langle U(x)|$

with the outcome distribution $(\mathscr{X}, \tau/d^2)$ *satisfying*

$$\frac{1}{d^2}\int_{\mathscr{X}} \mathrm{d}\tau(x) U(x)^{\otimes 2} \otimes (U(x)^{\otimes 2})^\dagger = \int_{\mathrm{PU}(d)} \mathrm{d}\mu(x) U(x)^{\otimes 2} \otimes (U(x)^{\otimes 2})^\dagger. \tag{6.11}$$

That is, if $\mathscr{X}$ *is finite, then* $(\mathscr{X}, \tau/d^2)$ *is a weighted unitary 2-design.*

Weighted unitary 2-designs thus define the class of optimal measurements on the output state for linear ancilla-assisted process tomography of unital quantum channels. Proposition 6.2 and corollary 6.1 summarize this main result of the paper. By corollary 5.3, these measurements exist in all dimensions. One particularly interesting type is that specified by a complete set of MUUBs. This choice allows us to perform optimal ancilla-assisted process tomography through a series of orthogonal measurements on the output state. In all cases, the optimal reconstruction formula for the output state is (equation (4.29) with $D = d^2$ and $\delta = (d^2-1)^2 + 1$)

$$\rho = (d^2-1)\int_{\mathscr{X}} \mathrm{d}p(x)|U(x)\rangle\langle U(x)| - \left(1 - \frac{2}{d^2}\right) I, \tag{6.12}$$

where $p(\mathscr{E}) = \mathrm{tr}[F(\mathscr{E})\rho] = \int_{\mathscr{E}} \mathrm{d}\tau(x)\langle U(x)|\rho|U(x)\rangle$ are the measurement outcome statistics. The corresponding unital channel follows from the Jamiołkowski isomorphism (equation (2.3)),

$$\mathcal{E} = (d^2-1)\int_{\mathscr{X}} \mathrm{d}p(x) U(x) \odot U(x)^\dagger - \left(d - \frac{2}{d}\right)\mathbf{I}. \tag{6.13}$$

Finally, consider the case $\mathscr{Q} = \mathscr{Q}^{\mathrm{gc}}$. Our condition for a tight rank-1 POVM (equation (4.28)) with $D = d^2$, $\delta = d^2(d^2-1)+1$ and $\mathbf{\Pi}_{\mathcal{F}} = \mathbf{\Pi}_{\mathscr{Q}^{\mathrm{gc}}}$ now becomes

$$\mathcal{F} = \int_{\mathscr{X}} \mathrm{d}\tau(x)|P(x))(P(x)| = |\lambda_0 \otimes \lambda_0)(\lambda_0 \otimes \lambda_0| + \frac{1}{d^2}\sum_{\substack{j>0 \\ k}} |\lambda_j \otimes \lambda_k)(\lambda_j \otimes \lambda_k|. \tag{6.14}$$

But following the exact same procedure as in the unital case we find that the types of tight rank-1 POVMs corresponding to $\mathscr{Q}^{\mathrm{gc}}$ must also satisfy equation (6.9), and thus, we again have $P(x) = |U(x)\rangle\langle U(x)|$, τ-almost everywhere, for some function $U : \mathscr{X} \to \mathrm{U}(d)$. In this case $\mathrm{Tr}(\mathcal{F}^2) = 2 - 1/d^2$ under equation (6.14), however, which would violate theorem 5.4. Our only conclusion can be that tight rank-1 POVMs with $\mathrm{supp}(\mathcal{F}) = \mathrm{span}(\mathscr{Q}^{\mathrm{gc}})$ do not exist. The lower bound on the error rate (equation (6.5)) still applies, but it is unattainable.

7. Conclusion

In this paper, we have shown that weighted unitary 2-designs specify optimal measurements on the system-ancilla output state for ancilla-assisted process tomography of unital quantum channels (corollary 6.1 and proposition 6.2). Although existence is known in all dimensions (corollary 5.3), it remains to construct specific examples of these designs with sizes as close as possible to the lower bound (theorem 5.5). Complete sets of MUUBs are known in dimensions $d = 2, 3, 5, 7, 11$, and form unweighted unitary 2-designs with sizes close to optimality. Each of these in fact specifies a minimal series of optimal orthogonal measurements (theorem 5.6). Weighted unitary 2-designs of smaller size exist in dimension 2, however (see table 1), and thus further reductions should be expected in higher dimensions. The optimization of the measurements used for ancilla-assisted process tomography of general quantum channels remains an open problem.

Acknowledgments

The author would like to thank Aidan Roy for helpful discussions. This work is supported by ARC and the State of Queensland.

Appendix. Quantum operations

Before describing quantum operations let us take a moment to set notation. Following Caves [54] (see also [55, 56]) we write a linear operator A in vector notation as $|A)$. The vector space of all such operators, $\mathrm{End}(\mathbb{C}^d) \cong \mathbb{C}^{d^2}$, equipped with the Hilbert–Schmidt inner product $(A|B) := \mathrm{tr}(A^\dagger B)$, is a Hilbert space, where we think of $(A|$ as an operator 'bra' and $|B)$ as an operator 'ket'. Addition and scalar multiplication of operator kets then follows that for operators, e.g. $a|A) + b|B) = |aA + bB)$. The usefulness of this notation becomes apparent when we consider linear maps on operators, i.e. superoperators. Given an orthonormal operator basis $\{E_k\}_{k=1}^{d^2} \subset \mathrm{End}(\mathbb{C}^d)$, $(E_j|E_k) = \delta_{jk}$, a superoperator $\mathcal{S} \in \mathrm{End}(\mathrm{End}(\mathbb{C}^d)) \cong \mathbb{C}^{d^4}$ may be written in two different ways,

$$\mathcal{S} = \sum_{j,k} s_{jk} E_j \odot {E_k}^\dagger = \sum_{j,k} s_{jk} |E_j)(E_k| \qquad (s \in \mathbb{C}^{d^2 \times d^2}). \tag{A.1}$$

The first representation illustrates the *ordinary* action of the superoperator,

$$\mathcal{S}(A) := \sum_{j,k} s_{jk} E_j A {E_k}^\dagger, \tag{A.2}$$

which amounts to inserting A into the location of the '$\odot$' symbol. The second reflects the *left–right* action,

$$\mathcal{S}|A) := \sum_{j,k} s_{jk} |E_j)(E_k|A) = \sum_{j,k} s_{jk} E_j \mathrm{tr}\big({E_k}^\dagger A\big), \tag{A.3}$$

where the superoperator acts on operators just like an operator on vectors. The identity superoperators relative to the ordinary and left–right actions are, respectively, $\mathcal{I} := I \odot I$ and $\mathbf{I} := \sum_k |E_k)(E_k|$. We also define $\mathrm{Tr}(\mathcal{S}) := \sum_k (E_k|\mathcal{S}|E_k)$ and $\|\mathcal{S}\| := \sqrt{\mathrm{Tr}(\mathcal{S}^\dagger \mathcal{S})}$. The latter is the Frobenius norm of $\mathcal{S}$ induced by its left–right action. Here $\mathcal{S}^\dagger$ is the left–right adjoint, i.e., $(A|\mathcal{S}^\dagger|B) := (B|\mathcal{S}|A)^*$, and $\mathcal{R}\mathcal{S}$ denotes the left–right composition of two superoperators $(\mathcal{R}\mathcal{S})|A) := \mathcal{R}|B)$ where $|B) = \mathcal{S}|A)$. Consult [54–56] for analogous concepts relative to the ordinary action.

The particular choice of operator basis $\{E_k = E_{k_1 k_2} := |k_1\rangle\langle k_2|\}_{k_1,k_2=1}^d$, where $\{|k\rangle\}_{k=1}^d$ is a fixed 'standard' basis for $\mathbb{C}^d$, defines the so-called *Jamiołkowski isomorphism* [13]. The matrix s in equation (A.1), now called the *process matrix*, then satisfies

$$s_{jk} = s_{j_1 j_2, k_1 k_2} = d(\langle j_1| \otimes \langle j_2|)[(\mathcal{S} \otimes \mathcal{I})(|I\rangle\langle I|)](|k_1\rangle \otimes |k_2\rangle), \tag{A.4}$$

where $|I\rangle := \sum_{k=1}^d |k\rangle \otimes |k\rangle/\sqrt{d}$. Note that $\mathrm{Tr}(\mathcal{S}^\dagger \mathcal{S}) = \mathrm{tr}(s^\dagger s)$ in general, and thus, the Hilbert–Schmidt superoperator distance between $\mathcal{S}$ and $\mathcal{R}$ can be rewritten as $\|\mathcal{S} - \mathcal{R}\| = \|s - r\|$, where r and $\mathcal{R}$ are related through equation (A.1). The upshot of the current choice of operator basis is that when $\mathcal{S}$ and $\mathcal{R}$ are quantum channels, as in this paper, then s and r specify standard-basis matrix elements of output quantum states with fixed input $|I\rangle\langle I|$.

A superoperator $\mathcal{S}$ is called *positive* if it maps positive operators to positive operators under its ordinary action. If, in addition, for any auxiliary system $\mathbb{C}^{d_a}$ we have $(\mathcal{S} \otimes \mathcal{I})(A) \geqslant 0$ whenever $A \geqslant 0$, $A \in \mathrm{End}(\mathbb{C}^d \otimes \mathbb{C}^{d_a})$, then $\mathcal{S}$ is called *completely positive*. Alternatively, $\mathcal{S}$ is completely positive if and only if $(A|\mathcal{S}|A) \geqslant 0$ for all $A \in \mathrm{End}(\mathbb{C}^d)$ [57]. That is, $\mathcal{S}$

is completely positive if and only if $\mathcal{S}^\dagger = \mathcal{S}$ and $\mathcal{S}$ has non-negative left–right eigenvalues. Diagonalizing, we see that $\mathcal{S}$ is completely positive if and only if it can be rewritten in an operator-sum form, called the *Kraus representation* [57–59],

$$\mathcal{S} = \sum_k B_k \odot {B_k}^\dagger = \sum_k |B_k)(B_k|, \tag{A.5}$$

where the operators $B_k \in \mathrm{End}(\mathbb{C}^d)$ are called *Kraus operators*. A superoperator $\mathcal{S}$ is said to be *trace nonincreasing* if $\mathrm{tr}[\mathcal{S}(A)] \leqslant \mathrm{tr}(A)$ for all A, and moreover, *trace preserving* if $\mathrm{tr}[\mathcal{S}(A)] = \mathrm{tr}(A)$ for all A. Thus the Kraus operators together satisfy $\sum_k {B_k}^\dagger B_k \leqslant I$ when $\mathcal{S}$ is trace nonincreasing and $\sum_k {B_k}^\dagger B_k = I$ when $\mathcal{S}$ is trace preserving.

A *quantum operation* is a superoperator-valued measure $\mathcal{E}[\cdot] : \mathfrak{B}(\mathscr{X}) \to \mathrm{End}(\mathrm{End}(\mathbb{C}^d))$ on an outcome set $\mathscr{X}$, which satisfies (1) $\mathcal{E}[\mathscr{S}]$ is completely positive and trace nonincreasing for all $\mathscr{S} \in \mathfrak{B}(\mathscr{X})$ with $\mathcal{E}[\emptyset] = 0$, (2) $\mathcal{E}[\bigcup_{k=1}^\infty \mathscr{S}_k] = \sum_{k=1}^\infty \mathcal{E}[\mathscr{S}_k]$ for any sequence of disjoint sets $\mathscr{S}_k \in \mathfrak{B}(\mathscr{X})$, and (3) $\mathcal{E}[\mathscr{X}]$ is trace preserving. In this paper, we always take $\mathfrak{B}(\mathscr{X})$ to be the Borel σ-algebra. Quantum operations can be *nonselective* (e.g. channels), in which case there is only one output $\rho' = \mathcal{E}(\rho) := \mathcal{E}[\mathscr{X}](\rho)$ for each input ρ, but are generally *selective* (e.g. measurements), in which case the output $\rho' = \mathcal{E}[\mathscr{S}](\rho)/p(\mathscr{S})$ occurs with probability $p(\mathscr{S}) = \mathrm{tr}(\mathcal{E}[\mathscr{S}](\rho))$. To make the connection to quantum measurements simply note that $F(\cdot) := \sum_k A_k(\cdot)^\dagger A_k(\cdot)$ is the POVM describing the outcome statistics of the measuring instrument $\mathcal{E}[\cdot] = \sum_k A_k(\cdot) \odot A_k(\cdot)^\dagger$.

References

[1] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[2] Paris M and Řeháček J (ed) 2004 *Quantum State Estimation* (Berlin: Springer)
[3] Leung D W 2003 Choi's proof as a recipe for quantum process tomography *J. Math. Phys.* **44** 528
[4] D'Ariano G M and Presti P Lo 2001 Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation *Phys. Rev. Lett.* **86** 4195
[5] Scott A J 2006 Tight informationally complete quantum measurements *J. Phys. A: Math. Gen.* **39** 13507
[6] Daubechies I, Grossmann A and Meyer Y 1986 Painless nonorthogonal expansions *J. Math. Phys.* **27** 1271
[7] Delsarte P, Goethals J M and Seidel J J 1977 Spherical codes and designs *Geom. Dedicata* **6** 363
[8] Neumaier A 1981 Combinatorial configurations in terms of distances *Dept. of Mathematics Memorandum 81–09* (Eindhoven University of Technology)
[9] Hoggar S G 1982 t-designs in projective spaces *Eur. J. Comb.* **3** 233
[10] Dankert C, Cleve R, Emerson J and Livine E Exact and approximate unitary 2-designs: constructions and applications *Preprint* quant-ph/0606161
[11] Gross D, Audenaert K and Eisert J 2007 Evenly distributed unitaries: on the structure of unitary designs *J. Math. Phys.* **48** 052104
[12] Landau L J and Streater R F 1993 On Birkoff's theorem for doubly stochastic completely positive maps of matrix algebras *Linear Algebra Appl.* **193** 107
[13] Jamiołkowski A 1972 Linear transformations which preserve trace and positive semidefiniteness of operators *Rep. Math. Phys.* **3** 275
[14] Busch P, Lahti P J and Mittelstaedt P 1996 *The Quantum Theory of Measurement* 2nd edn (Berlin: Springer)
[15] Prugovečki E 1977 Information-theoretic aspects of quantum measurement *Int. J. Theor. Phys.* **16** 321
[16] Busch P 1991 Informationally complete sets of physical quantities *Int. J. Theor. Phys.* **30** 1217
[17] Hradil Z 1997 Quantum-state estimation *Phys. Rev.* A **55** R1561
[18] Banaszek K, D'Ariano G M, Paris M G A and Sacchi M F 1999 Maximum-likelihood estimation of the density matrix *Phys. Rev.* A **61** 010304
[19] Jones K R W 1991 Principles of quantum inference *Ann. Phys.* **207** 140
[20] Bužek V, Derka R, Adam G and Knight P L 1998 Reconstruction of quantum states of spin systems: from quantum Bayesian inference to quantum tomography *Ann. Phys.* **266** 454
[21] Schack R, Brun T A and Caves C M 2001 Quantum Bayes rule *Phys. Rev.* A **64** 014305
[22] Tanaka F and Komaki F 2005 Bayesian predictive density operators for exchangeable quantum-statistical models *Phys. Rev.* A **71** 052323

[23] Blume-Kohout R Optimal, reliable estimation of quantum states *Preprint* quant-ph/0611080
[24] Christensen O 2003 *An Introduction to Frames and Riesz Bases* (Boston: Birkhäuser)
[25] Seymour P D and Zaslavsky T 1984 Averaging sets: a generalization of mean values and spherical designs *Adv. Math.* **52** 213
[26] Welch L R 1974 Lower bounds on the maximum cross correlation of signals *IEEE Trans. Inform. Theory* **20** 397
[27] Diaconis P and Shahshahani M 1994 On the eigenvalues of random matrices *J. Appl. Probab.* **31** 49
[28] Rains E M 1998 Increasing subsequences and the classical groups *Electron. J. Combin.* **5** R12
[29] Sagan B E 2001 *The Symmetric Group* (New York: Springer)
[30] Horn R T, Scott A J, Walgate J, Cleve R, Lvovsky A I and Sanders B C 2005 Classical and quantum fingerprinting with shared randomness and one-sided error *Quantum Inform. Comput.* **5** 258
[31] Collins B 2003 Moments and cumulants of polynomial random variables on unitary groups, the Itzykson–Zuber integral and free probability *Int. Math. Res. Notes* **17** 953
[32] Collins B and Śniady P 2006 Integration with respect to the Haar measure on unitary, orthogonal and symplectic group *Commun. Math. Phys.* **264** 773
[33] Levenshtein V 1998 On designs in compact metric spaces and a universal bound on their size *Discrete Math.* **192** 251
[34] Levenshtein V 1998 Universal bounds for codes and designs *Handbook of Coding Theory* ed V Pless and C W Huffman (Amsterdam: Elsevier) p 499
[35] Renes J M, Blume-Kohout R, Scott A J and Caves C M 2004 Symmetric informationally complete quantum measurements *J. Math. Phys.* **45** 2171
[36] Ivanović I D 1981 Geometrical description of quantal state determination *J. Phys. A: Math. Gen.* **14** 3241
[37] Wootters W K and Fields B D 1989 Optimal state-determination by mutually unbiased measurements *Ann. Phys.* **191** 363
[38] Roy A and Scott A J 2007 Weighted complex projective 2-designs from bases: optimal state determination by orthogonal measurements *J. Math. Phys.* **48** 072110
[39] Appleby D M 2005 Symmetric informationally complete-positive operator valued measures and the extended Clifford group *J. Math. Phys.* **46** 052107
[40] Flammia S T 2006 On SIC-POVMs in prime dimensions *J. Phys. A: Math. Gen.* **39** 13483
[41] Chau H F 2005 Unconditionally secure key distribution in higher dimensions by depolarization *IEEE Trans. Inform. Theory* **51** 1451
[42] Bannai E and Damerell R 1979 Tight spherical designs I *J. Math. Soc. Japan* **31** 199
[43] Bannai E and Damerell R 1980 Tight spherical designs II *J. London Math. Soc.* **21** 13
[44] Bannai E, Munemasa A and Venkov B 2004 The nonexistence of certain tight spherical designs *Algebra i Analiz* **16** 1 (Russian)
Bannai E, Munemasa A and Venkov B 2005 The nonexistence of certain tight spherical designs *St. Petersburg Math. J.* **16** 609 (Engl. Transl.)
[45] Reznick B 1995 Some constructions of spherical 5-designs *Linear. Algebra Appl.* **226-228** 163
[46] Sloane N J A, Hardin R H and Cara P 2003 Spherical designs in four dimensions *Proc. 2003 IEEE Information Theory Workshop, La Sorbonne, Paris, France, 31 March–4 April 2003* p 253
[47] Hardin R H and Sloane N J A 1994 Expressing $(a^2 + b^2 + c^2 + d^2)^3$ as a sum of 23 sixth powers *J. Combin. Theory* A **68** 481
[48] Harpe P de la, Pache C and Venkov B 2006 Construction of spherical cubature formulas using lattices *Algebra i Analiz* **18** 162 (Russian)
Harpe P de la, Pache C and Venkov B 2007 Construction of spherical cubature formulas using lattices *St. Petersburg Math. J.* **18** 119 (Engl. Transl.)
[49] Boyvalenkov P and Danev D 2001 Uniqueness of the 120-point spherical 11-design in four dimensions *Arch. Math.* **77** 360
[50] Andreev N N 2000 A minimal design of order 11 on the three-dimensional sphere *Mat. Zametki* **67** 489 (Russian)
Andreev N N 2000 A minimal design of order 11 on the three-dimensional sphere *Math. Notes* **67** 417 (Engl. Transl.)
[51] Yudin V A 1997 Lower bounds for spherical designs *Izv. Ross. Akad. Nauk Ser. Mat.* **61(3)** 213 (Russian)
Yudin V A 1997 Lower bounds for spherical designs *Izv. Math.* **61** 673 (Engl. Transl.)
[52] Salihov G N 1975 Cubature formulas for a hypersphere that are invariant with respect to the group of the regular 600-face *Dokl. Akad. Nauk SSSR* **223** 1098 (Russian)
Salihov G N 1975 Cubature formulas for a hypersphere that are invariant with respect to the group of the regular 600-face *Sov. Math. Dokl.* **16** 1046 (Engl. Transl.)
[53] Shamsiev É A 2006 On calculation of integrals over spherical domains *Ukrainian Math. J.* **58** 974
[54] Caves C M 1999 Quantum error correction and reversible operations *J. Supercond.* **12** 707

[55] Rungta P, Munro W J, Nemoto K, Deuar P, Milburn G J and Caves C M 2000 Qudit entanglement *Directions in Quantum Optics: A Collection of Papers Dedicated to the Memory of Dan Walls* ed H J Carmichael, R J Glauber and M O Scully (Berlin: Springer) p 149
[56] Rungta P, Buzek V, Caves C M, Hillery M and Milburn G J 2001 Universal state inversion and concurrence in arbitrary dimensions *Phys. Rev.* A **64** 042315
[57] Choi M-D 1975 Completely positive linear maps on complex matrices *Linear. Algebra Appl.* **10** 285
[58] Kraus K 1971 General state changes in quantum theory *Ann. Phys.* **64** 311
[59] Kraus K 1983 *States, Effects and Operations* (Berlin: Springer)