Commun. Math. Phys. 291, 813–843 (2009)
Digital Object Identifier (DOI) 10.1007/s00220-009-0890-5

Communications in
Mathematical
Physics

Distinguishability of Quantum States Under Restricted Families of Measurements with an Application to Quantum Data Hiding

William Matthews[1], **Stephanie Wehner**[2], **Andreas Winter**[1,3]

[1] Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K.
E-mail: will.x.matthews@googlemail.com, william.matthews@bris.ac.uk, a.j.winter@bris.ac.uk
[2] Institute for Quantum Information, Caltech, Pasadena, CA 91125, USA.
E-mail: wehner@caltech.edu
[3] Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, Singapore 117542, Singapore

Received: 31 October 2008 / Accepted: 4 June 2009
Published online: 13 August 2009 –

Abstract: We consider the problem of ambiguous discrimination of two quantum states when we are only allowed to perform a restricted set of measurements. Let the bias of a POVM be defined as the total variational distance between the outcome distributions for the two states to be distinguished. The performance of a set of measurements can then be defined as the ratio of the bias of this POVM and the largest bias achievable by any measurements. We first provide lower bounds on the performance of various POVMs acting on a single system such as the isotropic POVM, and spherical 2 and 4-designs, and show how these bounds can lead to certainty relations. Furthermore, we prove lower bounds for several interesting POVMs acting on multipartite systems, such as the set of local POVMS, POVMs which can be implemented using local operations and classical communication (LOCC), separable POVMs, and finally POVMs for which every bipartition results in a measurement having positive partial transpose (PPT). In particular, our results show that a scheme of Terhal et. al. for hiding data against local operations and classical communication [31] has the best possible dimensional dependence.

1. Introduction

Suppose we are given one of two possible quantum states ρ_0 and ρ_1 chosen with probabilities π_0 and π_1, respectively. The goal of ambiguous state discrimination is to output a guess ρ_b for the given state such that the average probability of error is minimized. To obtain this guess, we may thereby perform a measurement providing us with an outcome $b \in \{0, 1\}$. In this paper, we study the task of state discrimination when we are only allowed to perform a restricted set of measurements. To state our results, let us first explain the notion of a measurement (POVM) more formally. Consider a measurable space $(X, \mathcal{F})$, that is $\mathcal{F}$ is a σ-algebra of subsets of the set X, where we will identify $\mathcal{F}$ with the possible outcomes of the measurement. A POVM is a function $\mathrm{M} : \mathcal{F} \to \mathcal{B}^+(\mathcal{H})$ such that $\mathrm{M}(X) = \mathbb{1}$, where we use $\mathcal{B}^+(\mathcal{H})$ to denote the set of positive Hermitian operators on a finite dimensional Hilbert space $\mathcal{H}$. That is, $\mathrm{M}(A)$ is the measurement operator

associated with outcome event $A \in \mathcal{F}$. With regard to the problem of state discrimination, suppose that when performing the POVM M we guess that the state is ρ_1 if we observe outcome $A \in \mathcal{F}$, and ρ_0 otherwise. The probability of error can then be written as

$$P_{err} = \pi_0 \operatorname{Tr}(\mathrm{M}(A)\rho_0) + \pi_1 \operatorname{Tr}(\mathrm{M}(X \setminus A)\rho_1).$$

Let us now consider which outcomes $A \in \mathcal{F}$ we should associate with ρ_1 and ρ_0 respectively in order to minimize the probability of error. We will use $\mathcal{B}^{\mathrm{sa}}(\mathcal{H})$ to denote the space of Hermitian operators, and $\mathcal{D}(\mathcal{H}) = \{\rho \in \mathcal{B}^{+}(\mathcal{H}) \mid \operatorname{Tr}(\rho) = 1\}$ to denote the space of density operators on $\mathcal{H}$. Note that for any Hermitian operator $\xi \in \mathcal{B}^{\mathrm{sa}}(\mathcal{H})$, the function $\nu_{\mathrm{M}}[\xi] : \mathcal{F} \to \mathbb{R}$ defined by

$$\nu_{\mathrm{M}}[\xi](A) := \operatorname{Tr}(\mathrm{M}(A)\xi)$$

is a signed measure on $(X, \mathcal{F})$. It has a Hahn–Jordan decomposition [2]

$$\nu_{\mathrm{M}}[\xi] = \nu_{\mathrm{M}}[\xi]^{+} - \nu_{\mathrm{M}}[\xi]^{-},$$

where $\nu_{\mathrm{M}}[\xi]^{+}$ and $\nu_{\mathrm{M}}[\xi]^{-}$ are positive measures on $(X, \mathcal{F})$. For all $A \in \mathcal{F}$ these measures can be written as

$$\nu_{\mathrm{M}}[\xi]^{+}(A) := \nu_{\mathrm{M}}[\xi](A \cap X_{+}), \tag{1}$$

$$\nu_{\mathrm{M}}[\xi]^{-}(A) := \nu_{\mathrm{M}}[\xi](A \cap X_{-}), \tag{2}$$

where X_{+} and X_{-} are the positive and negative parts of the Hahn decomposition of X with respect to $\nu_{\mathrm{M}}[\xi]$. Note that we can rewrite the probability of error using the function $\nu_{\mathrm{M}}[\xi]$ as

$$P_{err} = \pi_1 - \operatorname{Tr}\mathrm{M}(A)(\pi_1\rho_1 - \pi_0\rho_0) = \pi_1 - \nu_{\mathrm{M}}[\xi](A) \quad \text{with } \xi = \pi_1\rho_1 - \pi_0\rho_0. \tag{3}$$

In light of the Hahn–Jordan decomposition, it is now clear that the smallest probability of error is attained by letting $A = X_{+}$ correspond to the guess ρ_1, yielding

$$P_{err} = \pi_1 - \nu_{\mathrm{M}}[\xi](X_{+}) = \pi_1 - \frac{1}{2}\left(\nu_{\mathrm{M}}[\xi](X_{+}) - \nu_{\mathrm{M}}[\xi](X_{-}) + \nu_{\mathrm{M}}[\xi](X)\right) \tag{4}$$

$$= \frac{1}{2}\left(1 - (\nu_{\mathrm{M}}[\xi](X_{+}) - \nu_{\mathrm{M}}[\xi](X_{-}))\right) = \frac{1}{2}(1 - \|\nu_{\mathrm{M}}[\xi]\|), \tag{5}$$

where

$$\|\nu_{\mathrm{M}}[\xi]\| := |\nu_{\mathrm{M}}[\xi](X_{+})| + |\nu_{\mathrm{M}}[\xi](X_{-})|$$

is the *total variation* of the signed measure $\nu_{\mathrm{M}}[\xi]$. Defining the *bias* to be $1 - 2P_{err}$, we have shown that the largest bias that can be attained, based on the outcomes of M, is $\|\nu_{\mathrm{M}}[\xi]\|$. Hence if we are only able to implement the POVMs in some set fixed set $\mathbf{M}$, the best bias that can be achieved is given by

$$\|\xi\|_{\mathbf{M}} := \sup_{\mathrm{M} \in \mathbf{M}} \|\nu_{\mathrm{M}}[\xi]\|. \tag{6}$$

When the measurable space $(X, \mathcal{F})$ has countable or finite $X = \{x_1, x_2, \ldots\}$ with $\mathcal{F}$ containing all subsets of X, the total variation of a signed measure ν is simply

$$\|\nu\| = \sum_{x \in X} |\nu(\{x\})|.$$

In this case, $\|\nu_{\mathrm{M}}[\xi]\|$ is just the ℓ_1 norm of the vector $(\operatorname{Tr} \mathrm{M}(x_1)\xi, \operatorname{Tr} \mathrm{M}(x_2)\xi, \ldots)$, and the operators $\mathrm{M}(x_i)$ are often called the elements of the POVM.

The Holevo–Helstrom theorem [23] tells us that when we are allowed to perform *any* POVM, the choice of measurement that maximises the bias is the two–outcome POVM with elements equal to the projectors onto the postive and negative eigenspaces of ξ. It is not hard to see that this achieves a bias equal to the trace norm of ξ, $\|\xi\|_1$. A natural indicator for the performance of a restricted set of POVMs $\mathbf{M}$ is thus given by the ratio

$$\frac{\|\xi\|_{\mathbf{M}}}{\|\xi\|_1}. \tag{7}$$

A. Results. We first show in Sect. 2 that (6) is a norm for every sufficiently rich set $\mathbf{M}$. We furthermore make a connection to general norms in vector spaces, showing that indeed any norm on trace class operators can be interpreted as a norm of the above type. We then turn to a number of particular examples, where we especially highlight the problem of determining bounds on the ratio (7). In Sect. 3, we investigate the particular case where $\mathbf{M}$ consists of only one (necessarily informationally complete) POVM, finding the best upper and lower bounds on the ratio (7) for any such measurement. These bounds are attained for the isotropic (unitary invariant) POVM. We also analyse the situation for POVMs originating from 2- and 4-designs. In Section 4, we look at the situation that the system under consideration is bi- or multipartite, and that the POVMs are restricted to classes respecting the partition: local measurements, with or without classical communication between the parties, and extensions of this class. The existence of data hiding [14,22,31] states yields bounds on the ratio (7) in one direction. Here, we show that in the bipartite case, these bounds are optimal up to a constant factor by analysing the tensor product of two isotropic local POVMs: it turns out that the resulting measurement attains almost the same bias. Hence, the hiding states of [31] are already (near) optimal in the sense that we cannot hope to construct states which are less well distinguishable under LOCC operations. Finally, we make a connection to Sanchez-Ruiz' "certainty relations" for mutually unbiased bases [29] in Sect. 5, which we show holds more generally for any 2-design POVM, and – even in a stronger form – for 4-designs. We also show how our results for bipartite systems imply a universal lower bound on the information accessible by LOCC from any pure state ensemble. Several appendices contain the proofs of more technical results in the main text.

2. First Observations on Norms and Dual Norms

Before turning to the essential observations that we will need later on, we first explain some basic concepts. We follow the terminology of Rockafellar [3] when referring to some elementary concepts from convex analysis. For norms $\|\cdot\|_a$ and $\|\cdot\|_b$ defined on a space V, we write $\|\cdot\|_a \geq \|\cdot\|_b$ if $\|x\|_a \geq \|x\|_b$ for all x in V. At the heart of the Helstrom-Holevo Theorem [23] on optimal state discrimination lies the duality between

the operator norm $\|\cdot\|$ and the trace norm $\|\cdot\|_1$: For operators α, A on a Hilbert space $\mathcal{H}$, these are dual to each other with

$$\|\alpha\|_1 = \sup_{\|B\|\leq 1} |\operatorname{Tr}(\alpha^\dagger B)|,$$
$$\|A\| = \sup_{\|\beta\|_1\leq 1} |\operatorname{Tr}(\beta^\dagger A)|.$$

In finite dimension, which we shall assume throughout this paper, the suprema are easily seen to be maxima. The duality persists when we restrict to Hermitian (self-adjoint) operators $\alpha = \alpha^\dagger$, $A = A^\dagger$:

$$\|\alpha\|_1 = \max_{B=B^\dagger,\ \|B\|\leq 1} \operatorname{Tr}(\alpha B),$$
$$\|A\| = \max_{\beta=\beta^\dagger,\ \|\beta\|_1\leq 1} \operatorname{Tr}(\beta A).$$

These equations are direct consequences of the singular value decomposition in the general, and of the spectral theorem in the Hermitian case.

The role of the Hilbert-Schmidt inner product,

$$\langle A, B\rangle := \operatorname{Tr} A^\dagger B$$

which makes the real vector space of Hermitian operators, $\mathcal{B}^{\mathrm{sa}}(\mathcal{H})$, a Euclidean space, becomes more evident in geometrical language by saying that the unit balls

$$B_1(\|\cdot\|_1) = \left\{\alpha = \alpha^\dagger : \|\alpha\|_1 \leq 1\right\},$$
$$B_1(\|\cdot\|) = \left\{A = A^\dagger : \|A\| \leq 1\right\},$$

are *polar* to each other. To explain this notion, note that the unit ball of any norm N on a finite dimensional real vector space,

$$K := B_1(N) = \{x : N(x) \leq 1\},$$

is a topologically compact, convex and symmetric set (i.e. $K = -K$), containing the origin 0 in its interior. Any such body K conversely determines a norm

$$\|x\|_{\check{K}} = \inf\left\{\frac{1}{t} : t > 0 \text{ and } tx \in K\right\},$$

and it is immediately verified that $K = B_1(\|\cdot\|_{\check{K}})$ and $N = \|\cdot\|_{\check{K}}$ (unconventionally, we write $\|\cdot\|_K$ for the norm with unit ball $\check{K}$ rather than that with unit ball K, as it simplifies the notation later). That is, norms and convex, compact, symmetric bodies of non–empty interior are equivalent descriptions. Now, the polar of K in a Euclidean vector space with inner product $\langle\cdot,\cdot\rangle$ is defined to be

$$\check{K} := \{y : \forall x \in K\ \langle x, y\rangle \leq 1\}.$$

It is easy to verify that if K is symmetric, convex and compact, and contains the origin in its interior, then $\check{K}$ has the same properties, and $\check{\check{K}} = K$.

By the above discussion, K is the unit ball of $\|\cdot\|_{\check{K}}$, $\check{K}$ is the unit ball of $\|\cdot\|_K$ and one has the important, but elementary, formulas

$$\|y\|_K = \max_{x\in K}\langle x, y\rangle,$$
$$\|x\|_{\check{K}} = \max_{y\in \check{K}}\langle x, y\rangle,$$

which are the abstract versions of the equations above. We are now ready to make a series of observations. First, we need to show that Eq. (6) really does constitute a norm for trace class operators, i.e. for operators with a finite, well-defined, trace.

First we note that, for any POVM M, $\|\nu_{\mathrm{M}}[\xi]\|$ is a seminorm on trace class operators ξ, which we give the shorthand $\|\xi\|_{\mathrm{M}} := \|\nu_{\mathrm{M}}[\xi]\|$. For sets of POVMs $\mathbf{M}$ have defined $\|\cdot\|_{\mathbf{M}} = \sup_{\mathrm{M}\in\mathbf{M}}\|\cdot\|_{\mathrm{M}}$ which, being a supremum over seminorms, is also a seminorm. Clearly, $\|\cdot\|_{\mathbf{M}}$ is a norm iff for all $\xi \neq 0$, there is a POVM $\mathrm{M} \in \mathbf{M}$ such that $\|\nu_{\mathrm{M}}[\xi]\| > 0$. We call a set of POVMs which satisfies this property "*informationally complete*".

It is often said that a set of POVMs is informationally complete iff knowledge of the statistics of the POVMs in the set is sufficient to reconstruct any unknown state (operationally, we think of having an unlimited number of copies of the state on which we can perform the measurements). It is not hard to see that this is true of a set $\mathbf{M}$ iff $\mathrm{span}\{\mathrm{M}(E) : \mathrm{M} \in \mathbf{M}, E \in F_{\mathrm{M}}\} =: S = \mathcal{B}(\mathcal{H})$. If there is a ξ such that $\|\xi\|_{\mathbf{M}} = 0$, then we must have $\mathrm{Tr}\,\mathrm{M}(E)\xi = 0$ for all POVMs and events, so $S \neq \mathcal{B}(\mathcal{H})$. Conversely, if $S \neq \mathcal{B}(\mathcal{H})$, then there is an operator ξ in the orthogonal complement of S and $\|\xi\|_{\mathbf{M}} = 0$. Therefore, the two definitions of informational completeness coincide.

We now show that we can restrict ourselves to POVMs with 2 outcomes. Intuitively, since we decide between two options (e.g. ρ and σ above), we can group the outcomes of each POVM in two. It is then not difficult to verify that

Definition 1. *For any separating set* $\mathbf{M}$ *of POVMs we define the set of two–outcome POVMs*

$$\mathbf{M}_2 := \bigcup_{\mathrm{M}\in\mathbf{M}} \{(\mathrm{M}(A), \mathbb{1} - \mathrm{M}(A)) : \forall\, A \in \mathcal{F}_{\mathrm{M}}\},$$

where $\mathcal{F}_{\mathrm{M}}$ *is the set of measurable subsets of outcomes for* M*, satisfies* $\|\cdot\|_{\mathbf{M}} = \|\cdot\|_{\mathbf{M}_2}$ *and we define*

$$\mathbb{M} := \mathrm{cl}\,\mathrm{conv}\,\{2E - \mathbb{1} : (E, \mathbb{1} - E) \in \mathbf{M}_2\} = \mathrm{cl}\,\mathrm{conv}\,\{2\mathrm{M}(A) - \mathbb{1} : A \in \mathcal{F}_{\mathrm{M}}\},$$

where cl conv S *denotes the closure of the convex hull of* S.

Lemma 2. $\mathbb{M}$ *is a compact symmetric convex body, contained in the operator interval* $[-\mathbb{1}; \mathbb{1}] = \{X : -\mathbb{1} \leq X \leq \mathbb{1}\}$ *and containing* $\pm\mathbb{1}$*, and has a non–empty interior, such that*

$$\|\xi\|_{\mathbf{M}} = \max_{E\in\mathbb{M}} |\,\mathrm{Tr}(\xi E)| =: \|\xi\|_{\mathbb{M}}.$$

Proof. From the discussion in the Introduction, it is clear that for a particular choice of ξ, the bias for any POVM $\mathrm{M} \in \mathbf{M}$ is equal to the bias for the two–outcome POVM $(\mathrm{M}(X_+), \mathrm{M}(X_-))$. □

Note that $\mathbb{M}$ has a non-empty interior (and then contains the origin in its interior) if and only if the collection $\mathbf{M}$ is informationally complete, which is the case if and only if $\mathbf{M}_2$ is informationally complete. Mathematically the information-completeness is expressed by $\mathbb{M}$, spanning the whole operator space. Furthermore, note that from our discussion above we have that

Remark 3. The symmetric convex body $\mathbb{M}$ defines two norms, one on the observables and effects, the other on the trace class operators, via

$$\|E\|_{\check{\mathbb{M}}} = \inf\left\{\frac{1}{t} : t > 0 \text{ and } tM \in \mathbb{M}\right\}, \tag{8}$$

$$\|\xi\|_{\mathbb{M}} = \max_{E\in\mathbb{M}} \operatorname{Tr}(\xi E). \tag{9}$$

The first has exactly $\mathbb{M}$ as its unit ball, the second has as its unit ball the *polar* of $\mathbb{M}$, i.e.

$$\check{\mathbb{M}} = \{\xi : \forall M \in \mathbb{M}\ \operatorname{Tr}(\xi M) \leq 1\}.$$

The norm $\|\cdot\|_{\mathbb{M}} (= \|\cdot\|_{\mathbf{M}})$ is dual to $\|\cdot\|_{\check{\mathbb{M}}}$:

$$\|\xi\|_{\mathbb{M}} = \max\left\{\operatorname{Tr}(\xi E) : \|E\|_{\check{\mathbb{M}}} \leq 1\right\},$$
$$\|E\|_{\check{\mathbb{M}}} = \max\left\{\operatorname{Tr}(\xi E) : \|\xi\|_{\mathbb{M}} \leq 1\right\}.$$

Putting everything together, we can now see that

Theorem 4. *The norms $\|\cdot\|_{\mathbf{M}}$ associated to sets of POVMs are in one-to-one correspondence with full-dimensional symmetric compact convex bodies $\pm\mathbb{1} \in \mathbb{M} \subseteq [-\mathbb{1}; \mathbb{1}]$.*

As a consequence, any *norm $|||\cdot||| \leq \|\cdot\|_1$ can be written as $|||\cdot||| = \|\cdot\|_{\mathbf{M}}$ for some set of POVMs.*

Proof. First, starting with a set of POVMs $\mathbf{M}$ defining norms $\|\cdot\|_{\mathbf{M}}$, Lemma 2 describes how to construct $\mathbb{M}$, such that $\|\cdot\|_{\mathbf{M}} = \|\cdot\|_{\mathbb{M}}$.

Conversely, starting with a full-dimensional symmetric compact convex body $\mathbb{M} \subseteq [-\mathbb{1}; \mathbb{1}]$, we can construct a set of POVMs $\mathbf{M} = \{(M, \mathbb{1} - M) : M \in \mathbb{M} \text{ and } M \geq 0\}$ for which $\|\cdot\|_{\mathbf{M}} = \|\cdot\|_{\mathbb{M}}$. □

We formalise the connection with the state discrimination problem in the following theorem.

Theorem 5. *Let $\mathbf{M}$ be a set of POVMs on a given Hilbert space, and let $\mathbf{M}_2$ and $\mathbb{M}$ be defined as above. For any two states ρ and σ, consider the minimum error probability $P_E^{\mathbf{M}}$ of discriminating between these (a priori equiprobable states). Then,*

$$P_E^{\mathbf{M}} = \inf_{(M,\mathbb{1}-M)\in\mathbf{M}_2} \frac{1}{2} - \frac{1}{2}|\operatorname{Tr}((\rho-\sigma)M)| = \frac{1}{2} - \frac{1}{4}\|\rho-\sigma\|_{\mathbb{M}}.$$

That is, $\frac{1}{2}\|\rho-\sigma\|_{\mathbb{M}}$ is the bias achievable in discriminating ρ from σ when only measurements in $\mathbf{M}$ are allowed.

In finite dimension, which is the case we stick to in this paper, the operators also form a finite-dimensional space, and all these norms are "equivalent" in the sense that there are $\lambda', \mu' > 0$ such that

$$\lambda'\|\cdot\|_1 \leq \|\cdot\|_{\mathbb{M}} \leq \mu'\|\cdot\|_1. \tag{10}$$

By using the above correspondences and dualities, we see that this is equivalent to

$$\lambda'[-\mathbb{1};\mathbb{1}] \subseteq \mathbb{M} \subseteq \mu'[-\mathbb{1};\mathbb{1}]. \tag{11}$$

We will use $\lambda_1(\mathbb{M})$ ($\mu_1(\mathbb{M})$) to denote the largest λ' (smallest μ') in these equations. The numbers λ_1 and μ_1 are called the *constants of domination* of the norm $\|\cdot\|_{\mathbb{M}}$ (with respect to $\|\cdot\|_1$). In the following, our goal is to bound these constants of domination for various interesting classes of POVMs. These constants are especially interesting, since we know from Theorem 5 that they allow us to bound the bias that we can achieve when trying to distinguish two states ρ and σ with a restricted set of measurements.

Note that $\mu_1(\mathbb{M})$ is trivially 1 since for $\rho \geq 0$, $\|\rho\|_{\mathbb{M}} = \|\rho\|_1 = \mathrm{Tr}(\rho)$. Thus, we are motivated to restrict to *traceless* operators in Eq. (10). This is also the setting for which bounds on the constants of domination give us a bound on the bias of distinguishing two a priori equiprobable states ρ and σ. Let $\lambda(\mathbb{M})$ and $\mu(\mathbb{M})$ be the largest and smallest numbers λ' and μ', respectively, such that

$$\forall \xi \text{ with } \mathrm{Tr}(\xi) = 0 \quad \lambda\|\xi\|_1 \leq \|\xi\|_{\mathbb{M}} \leq \mu\|\xi\|_1. \tag{12}$$

Equivalently, in the dual picture we have to go to the quotient modulo multiples of the identity, $\mathbb{R}\mathbb{1}$:

$$\lambda[-\mathbb{1};\mathbb{1}]/_{\mathbb{R}\mathbb{1}} \subseteq \mathbb{M}/_{\mathbb{R}\mathbb{1}} \subseteq \mu[-\mathbb{1};\mathbb{1}]/_{\mathbb{R}\mathbb{1}}, \tag{13}$$

where, for a set of operators X, $X/_{\mathbb{R}\mathbb{1}} = \{x - \mathbb{1}\,\mathrm{Tr}\,x/\,\mathrm{Tr}\,\mathbb{1} : x \in X\}$. The following lemma characterizes λ_1 (μ_1) and λ (μ), and their respective relations.

Lemma 6. *For a set* $\mathbf{M}$ *of POVMs with associated convex body* $\mathbb{M}$*, the constants of domination can be expressed as the solutions of the following optimisation problems:*

$$\frac{1}{2}\lambda(\mathbb{M}) \leq \lambda_1(\mathbb{M}) = \inf_{\|\xi\|_1=1} \sup_{\mathrm{M}\in\mathbf{M}} \|\xi\|_{\mathrm{M}} \leq \inf_{\substack{\|\xi\|_1=1 \\ \mathrm{Tr}(\xi)=0}} \sup_{\mathrm{M}\in\mathbf{M}} \|\xi\|_{\mathrm{M}} = \lambda(\mathbb{M}),$$

$$1 = \mu_1(\mathbb{M}) = \sup_{\|\xi\|_1=1} \sup_{\mathrm{M}\in\mathbf{M}} \|\xi\|_{\mathrm{M}} \geq \sup_{\substack{\|\xi\|_1=1 \\ \mathrm{Tr}(\xi)=0}} \sup_{\mathrm{M}\in\mathbf{M}} \|\xi\|_{\mathrm{M}} = \mu(\mathbb{M}).$$

Here, for the purpose of λ *and* μ*,* ξ *may be thought of as* $\xi = \frac{1}{2}(\rho - \sigma)$ *for orthogonal states* ρ*,* σ*.*

Proof. The optimisation problems are an immediate consequence of the definitions, and we already argued that $\mu_1(\mathbb{M}) = 1$. To lower bound $\lambda_1(\mathbb{M})$ we proceed as follows: Given any ξ of trace norm 1, we can write it as

$$\xi = (1-p)\rho - p\sigma = (1-p)(\rho-\sigma) + (1-2p)\sigma = 2(1-p)\xi_0 + (1-2p)\sigma,$$

with orthogonal states ρ and σ, and $\xi_0 = \frac{1}{2}(\rho - \sigma)$. W.l.o.g. $0 \leq p \leq 1/2$, otherwise use $-\xi$. Now let $X_0 \in \mathbb{M}$ be optimal for ξ_0, i.e. $\|\xi_0\|_{\mathbb{M}} = \mathrm{Tr}(\xi_0 X_0)$, and test ξ with $X = (\mathbb{1} + X_0)/2 \in \mathbb{M}$. Note $X \geq 0$, so

$$\begin{aligned}
\|\xi\|_{\mathbb{M}} \geq \mathrm{Tr}(\xi X) &= 2(1-p)\,\mathrm{Tr}(\xi_0 X) + (1-2p)\,\mathrm{Tr}(\sigma X) \\
&= (1-p)\,\mathrm{Tr}(\xi_0 X_0) + \frac{1-2p}{2}\,\mathrm{Tr}(\sigma X) \\
&\geq \frac{1}{2}\,\mathrm{Tr}(\xi_0 X_0) = \frac{1}{2}\|\xi_0\|_{\mathbb{M}} \geq \frac{1}{2}\lambda(\mathbb{M}),
\end{aligned}$$

concluding the proof. □

What is the relation of the constants of domination for different sets $\mathbb{M}$ and $\mathbb{M}'$? Clearly, if $\mathbb{M} \subseteq \mathbb{M}'$, then $\lambda(\mathbb{M}) \leq \lambda(\mathbb{M}')$ and $\mu(\mathbb{M}) \leq \mu(\mathbb{M}')$. More interesting relations are obtained by using the convex structure.

For this purpose we look at convex combinations of POVMs in the sense of direct sums as follows. Let $\mathrm{M} : \mathcal{F} \to \mathcal{B}^{\mathrm{sa}}(\mathcal{H})$ and $\mathrm{M} : \mathcal{G} \to \mathcal{B}^{\mathrm{sa}}(\mathcal{H})$ be two POVMs for measurable spaces $(X, \mathcal{F})$ and $(Y, \mathcal{G})$. Let $(X \cup Y, \mathcal{K})$ be the direct sum of $(X, \mathcal{F})$ and $(Y, \mathcal{G})$, i.e. $\mathcal{K} = \{A : A \cap X \in \mathcal{F} \text{ and } A \cap Y \in \mathcal{G},\ \forall A \in \mathcal{P}(X \cup Y)\}$ (where $\mathcal{P}$ denotes powerset).

For $p \in [0, 1]$, define the direct convex combination $(1 - p)\mathrm{M} \oplus p\mathrm{N} : X \oplus Y \to \mathcal{B}^{\mathrm{sa}}(\mathcal{H})$ by specifying that, for all $A \in \mathcal{K}$,

$$((1 - p)\mathrm{M} \oplus p\mathrm{N})(A) = (1 - p)\mathrm{M}(A \cap X) + p\mathrm{N}(A \cap Y).$$

If we have two sets of POVMs, $\mathbf{M_1}$ and $\mathbf{M_2}$, then their direct sum convex combination is defined naturally as

$$(1 - p)\mathbf{M_1} \oplus p\mathbf{M_2} := \{(1 - p)\mathrm{M}_1 \oplus p\mathrm{M}_2 : \forall \mathrm{M}_1 \in \mathbf{M_1},\ \mathrm{M}_2 \in \mathbf{M_2}\}.$$

More generally, we can look at convex combinations of any finite or even countable number of POVMs and sets of POVMs. These constructions have a straightforward operational interpretation: implementing $p(E_k)_k \oplus (1 - p)(F_\ell)_\ell$ means tossing a biased coin, with p being the probability of heads, then measuring (E_k) if heads showed, and (F_ℓ) for tails. The coin toss is part of the measurement result.

Lemma 7. *Let $\mathbf{M}_i$ be sets of POVMs and $p_i \geq 0$ probabilities, and $\mathbf{R} = \bigoplus_i p_i \mathbf{M}_i$. Denote the corresponding convex bodies of operators $\mathbb{M}_i$ and $\mathbb{R}$. Then,*

$$\mathbb{R} = \sum_i p_i \mathbb{M}_i,$$

and consequently

$$\lambda(\mathbb{R}) \geq \sum_i p_i \lambda(\mathbb{M}_i), \quad \mu(\mathbb{R}) \leq \sum_i p_i \mu(\mathbb{M}_i).$$

Proof. The first relation is by inspection. For the inequalities, note that since we have

$$\lambda(\mathbb{M}_i)[-\mathbb{1}; \mathbb{1}]/_{\mathbb{R}\mathbb{1}} \subseteq \mathbb{M}_i/_{\mathbb{R}\mathbb{1}} \subseteq \mu(\mathbb{M}_i)[-\mathbb{1}; \mathbb{1}]/_{\mathbb{R}\mathbb{1}},$$

we clearly get

$$\sum_i p_i \lambda(\mathbb{M}_i)[-\mathbb{1}; \mathbb{1}]/_{\mathbb{R}\mathbb{1}} \subseteq \sum_i p_i \mathbb{M}_i/_{\mathbb{R}\mathbb{1}} \subseteq \sum_i p_i \mu(\mathbb{M}_i)[-\mathbb{1}; \mathbb{1}]/_{\mathbb{R}\mathbb{1}}.$$

□

In particular, since $[-\mathbb{1}; \mathbb{1}]$ is invariant under unitary conjugation, i.e. $U[-\mathbb{1}; \mathbb{1}]U^\dagger = [-\mathbb{1}; \mathbb{1}]$, the constants of domination also have this invariance and so we obtain immediately

Proposition 8. *For a probability measure $\mathrm{d}p(U)$ on the unitary group on $\mathcal{H}$, and any symmetric, convex body with non–empty interior, $\pm\mathbb{1} \in \mathbb{M} \subseteq [-\mathbb{1}; \mathbb{1}]$,*

$$\lambda\left(\int \mathrm{d}p(U) U\mathbb{M}U^\dagger\right) \geq \lambda(\mathbb{M}),$$
$$\mu\left(\int \mathrm{d}p(U) U\mathbb{M}U^\dagger\right) \leq \mu(\mathbb{M}).$$

In other words: symmetrisation makes $\mathbb{M}$ "look more like $[-\mathbb{1}; \mathbb{1}]$".

3. Single POVMs

Let us look now at the constants of domination λ and μ in the case that $\mathbf{M}$ consists of a single, informationally complete POVM M. We denote the constants of domination $\lambda(\mathrm{M})$ and $\mu(\mathrm{M})$.

A. Isotropic POVM. Given a D dimensional Hilbert space $\mathcal{H}$, let X denote the sphere of normalised pure states in $\mathcal{H}$ and let $\mathcal{F}$ be the Borel measurable subsets of X. We define the isotropic POVM $\mathrm{M_U} : \mathcal{F} \to \mathcal{B}^+(\mathcal{H})$ by

$$\mathrm{M_U}(A) = D \int_A |\psi\rangle\langle\psi| \mathrm{d}\psi$$

with $\mathrm{d}\psi$ the unitarily invariant probability measure on $(X, \mathcal{F})$.

If we take any POVM $\mathrm{M} : (X, \mathcal{F}) \to \mathcal{B}^+(\mathcal{H})$ we can construct a new POVM M' which corresponds to a measurement where a random unitary is drawn according to the Haar measure, and recorded, and then M is measured. M' takes outcomes in the product measure space $(X, \mathcal{F}) \times (\mathrm{SU}(D), \mathcal{G})$ (where $\mathcal{G}$ is the set of Borel measurable subsets of the Lie group $\mathrm{SU}(D)$). M' is defined on 'rectangles' $A \times B$, $A \in \mathcal{F}$, $B \in \mathcal{G}$ by

$$\mathrm{M}'(A \times B) = \int_B U \mathrm{M}(A) U^\dagger \mathrm{d}p(U),$$

where p denotes the Haar measure on $(\mathrm{SU}(D), \mathcal{G})$. Now, the total variation of the signed measure $\nu_{\mathrm{M}'}[\xi]$ is

$$\|\nu_{\mathrm{M}'}[\xi]\| = \sup_{(A_i)} \sum_i (\mathrm{Tr}\, \mathrm{M}(A_i)) D \int_{SU(D)} \left| \mathrm{Tr}\left(U \frac{\mathrm{M}(A_i)}{\mathrm{Tr}\,(\mathrm{M}(A_i))} U^\dagger \xi \right) \right| \mathrm{d}p(U),$$

where the supremum is over finite partitions of X into measurable sets A_i. For any $\mathrm{M}(A)$, $\frac{\mathrm{M}(A)}{\mathrm{Tr}\,\mathrm{M}(A)}$ has a spectral decomposition $\sum_{j=1}^D p_j |\psi_j\rangle\langle\psi_j|$, where $\sum_{j=1}^D p_j = 1$ and $p_j \geq 0$. Therefore,

$$\begin{aligned}
&D \int_{\mathrm{SU}(D)} \left| \mathrm{Tr}\left(U \frac{\mathrm{M}(A)}{\mathrm{Tr}\,\mathrm{M}(A)} U^\dagger \xi \right) \right| \mathrm{d}p(U) \\
&= D \int_{\mathrm{SU}(D)} \left| \mathrm{Tr}\left(U \sum_{j=1}^D p_j |\psi_j\rangle\langle\psi_j| U^\dagger \xi \right) \right| \mathrm{d}p(U) \qquad (14) \\
&= D \int_{\mathrm{SU}(D)} \left| \mathrm{Tr}\left(U|0\rangle\langle 0|U^\dagger \sum_{j=1}^D p_j V_j \xi V_j^\dagger \right) \right| \mathrm{d}p(U) \qquad (15) \\
&\leq D \int_{\mathrm{SU}(D)} \left| \mathrm{Tr}\left(U|0\rangle\langle 0|U^\dagger \xi \right) \right| \mathrm{d}p(U) = \|\xi\|_{\mathrm{M_U}}. \qquad (16)
\end{aligned}$$

Given this, and the fact that $\sum_i (\mathrm{Tr}\, \mathrm{M}(A_i)) = 1$ for any partition (A_i), $\|\nu_{\mathrm{M}'}[\xi]\|$ can be no larger than $\|\xi\|_{\mathrm{M_U}}$. This bound is attained, for example, whenever the POVM is 'rank–one' in the following sense:

Definition 9. *Call a POVM* $\mathrm{M} : \mathcal{F} \to \mathcal{B}^+(\mathcal{H})$ *with outcomes in the measurable space* $(X, \mathcal{F})$ *'rank–one' if there is a countable partition* (A_i) *of* X *into* $A_i \in \mathcal{F}$ *such that* $\operatorname{rank} \mathrm{M}(A_i) = 1$ *for all* A_i.

As a consequence of Proposition 8, we arrive at the following theorem.

Theorem 10. *The supremum of* $\lambda(\mathrm{M})$ *over all single POVMs* M *in dimension* D *is attained by the* isotropic *POVM* M_{U}. *In addition the infimum of* $\mu(\mathrm{M})$ *over all* rank–one *POVMs is* $\mu(\mathrm{M}_{\mathrm{U}})$.

$$\lambda(\mathrm{M}_{\mathrm{U}}) = \min_{\substack{1 \le a \le D/2 \\ b = D-a}} 1 - \frac{1}{D} \sum_{\substack{k=0,\ldots a-1 \\ \ell=0,\ldots b-1}} \left(\frac{a}{D}\right)^k \left(\frac{b}{D}\right)^\ell \binom{k+\ell}{k} = \sqrt{\frac{1}{D}}\left(\sqrt{\frac{2}{\pi}} \pm o(1)\right), \tag{17}$$

$$\mu(\mathrm{M}_{\mathrm{U}}) = \frac{1}{2}. \tag{18}$$

Proof. Since randomising a POVM over unitary transformations (which we record) cannot decrease λ we can, without loss of generality, perform the symmetrization over the Haar measure described above without decreasing λ. From our discussion of such symmetrized POVMs it is clear that none has a larger value of λ than the isotropic POVM. The statement about μ for rank–one POVMs follows from the fact that the symmetrized version of such a POVM has the same bias as the isotropic POVM.

For the constants of domination for M_{U}, first note that for any operator ξ,

$$\|\nu_{\mathrm{M}_{\mathrm{U}}}[\xi]\| = D \int \mathrm{d}\psi \, |\operatorname{Tr}(|\psi\rangle\langle\psi|\xi)|. \tag{19}$$

Note that since ξ is Hermitian, we may again take $\xi = (1-p)\rho - p\sigma$ for orthogonal operators ρ and σ. For Eq. (17), we then have by the unitary invariance of the uniform POVM and the triangle inequality, $\lambda(\mathrm{M}_{\mathrm{U}})$ is attained as $\|\xi\|_{\mathrm{M}_{\mathrm{U}}}$ for an operator of the form

$$\xi = \frac{1}{2a} P - \frac{1}{2b} Q, \quad \text{with a projector } P \text{ of rank } a, \text{ and } Q = \mathbb{1} - P, \; b = D - a,$$

where we may even take P to be the projector onto the subspace spanned by the first a computational basis vectors (again invoking unitary invariance). For this choice of operator, according to Eq. (19), and letting $p := a/D$,

$$\begin{aligned} \|\xi\|_{\mathrm{M}_{\mathrm{U}}} &= D \int \mathrm{d}\psi \left| \frac{1}{2a} \sum_{j=1}^{a} |\psi_j|^2 - \frac{1}{2b} \sum_{j=a+1}^{D} |\psi_j|^2 \right| \\ &= 1 - \frac{1}{D} \sum_{\substack{k=0,\ldots a-1 \\ \ell=0,\ldots b-1}} p^k (1-p)^\ell \binom{k+\ell}{k}, \end{aligned} \tag{20}$$

by Lemma 24 in Appendix B. It is quite natural to conjecture that the minimal choice of ranks is $a = \lfloor D/2 \rfloor$ and $b = \lceil D/2 \rceil$. In this case we have

$$\begin{aligned} \|\xi\|_{\mathrm{M}_{\mathrm{U}}} &= 1 - \frac{1}{D} \sum_{\substack{k=0,\ldots \lfloor D/2 \rfloor - 1 \\ \ell=0,\ldots \lceil D/2 \rceil - 1}} \left(\frac{\lfloor D/2 \rfloor}{D}\right)^k \left(\frac{\lceil D/2 \rceil}{D}\right)^\ell \binom{k+\ell}{k} \\ &= \sqrt{\frac{2}{\pi D}} \pm O\left(\frac{1}{D}\right), \end{aligned} \tag{21}$$

for large D. The analysis of the asymptotics is elementary but lengthy, and is here restricted to a few hints: We lose only terms of order $O(1/D)$ by focusing on even D, for which the formula evaluates to

$$\lambda(\mathrm{M_U}) = 1 - \frac{1}{D}\sum_{k,\ell=0}^{D/2-1} 2^{-k-\ell}\binom{k+\ell}{k} = \frac{1}{D}\sum_{k=0}^{D/2-1} 2^{-2k}\binom{2k}{k},$$

where we have used the following identity from Lemma 25, proved by induction on k:

$$\sum_{\ell=0}^{k} 2^{-k-\ell}\binom{k+\ell}{\ell} = 1. \tag{22}$$

Then a simple application of Stirling's formula (with explicit error bounds) yields Eq. (21).

However, since we have not been able to prove that the minimum value of the expression 20 occurs for this choice of ranks, we instead follow a different route: From the proof of Lemma 24 in Appendix B, we observe that for general a and b,

$$\|\xi\|_{\mathrm{M_U}} = \mathbb{E}\left|\frac{1}{2a}\sum_{j=1}^{a} X_j - \frac{1}{2b}\sum_{j=a+1}^{d} X_j\right|,$$

with independent $X_j \geq 0$, each distributed according to a rescaled χ^2_2 law. By definition, their expectation and variance are $\mathbb{E}X_j = 1$ and $\mathrm{Var}\, X_j = 1$, respectively (also, all higher moments are finite). Thus, by the central limit theorem,

$$\frac{1}{2a}\sum_{j=1}^{a} X_j \approx Y_0 \sim \mathcal{N}\left(1, \frac{1}{4a}\right), \quad \frac{1}{2b}\sum_{j=a+1}^{d} X_j \approx Y_1 \sim \mathcal{N}\left(1, \frac{1}{4b}\right),$$

where Y_0 and Y_1 are normal distributed with means μ and variance ν as indicated by $\mathcal{N}(\mu, \nu)$, and the approximation signs indicate convergence in probability as $a, b \to \infty$. (Note that since the third moment of the X_j is finite, this convergence is uniform in a and b, thanks to the Berry-Esséen theorem which bounds the rate of convergence in the central limit theorem – see e.g. [13].)

Since $Y_0 - Y_1 =: Z \sim \mathcal{N}\left(0, \frac{1}{4a} + \frac{1}{4b}\right)$, we obtain asymptotically

$$\|\xi\|_{\mathrm{M_U}} \sim \mathbb{E}|Z| = \sqrt{\frac{1}{4a} + \frac{1}{4b}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{d}x|x|e^{-x^2/2} = \sqrt{\frac{2}{\pi}}\sqrt{\frac{1}{4a} + \frac{1}{4b}},$$

which is minimized for $a = b = D/2$, yielding $\lambda(\mathrm{M_U}) \sim \sqrt{\frac{2}{\pi D}}$, as advertised.

For Eq. (18), note that by the triangle inequality, $\mu(\mathrm{M})$ of any POVM M is attained for an extremal traceless ξ such that $\|\xi\|_1 = 1$. These are easily seen to be of the form $\xi = \frac{1}{2}|\phi_1\rangle\langle\phi_1| - \frac{1}{2}|\phi_2\rangle\langle\phi_2|$ for orthogonal pure state vectors $|\phi_1\rangle, |\phi_2\rangle$. By unitary invariance of the uniform POVM, any such ξ will in fact yield the same value, so we may take $\xi = \frac{1}{2}|1\rangle\langle 1| - \frac{1}{2}|2\rangle\langle 2|$, so that by Eq. (19),

$$\mu(\mathrm{M_U}) = \|\xi\|_{\mathrm{M_U}} = \frac{D}{2}\int \mathrm{d}\psi \left||\psi_1|^2 - |\psi_2|^2\right| = \frac{1}{2},$$

once more by Lemma 24 in Appendix B, applied with $a = b = 1$. □

Note that in terms of the bias the above translates to

$$\frac{1}{2}\|\rho-\sigma\|_1 \geq \|\rho-\sigma\|_{\mathrm{M}} \geq \sqrt{\frac{1}{D}}\left(\sqrt{\frac{2}{\pi}}-o(1)\right)\|\rho-\sigma\|_1.$$

B. Almost optimal performance of 4*-designs.* The results of the previous section provide the motivation to look at POVMs made from t-designs, as these are structures approximating the full random POVM better and better as $t \to \infty$. We thus intuitively expect to obtain a similar value for λ as we obtained for the random POVM for larger t.

On the k^{th} tensor power $\mathcal{H}^{\otimes k}$ of a Hilbert space $\mathcal{H}$, there is a natural unitary representation of the permutation group of order k, S_k, which permutes the k tensor factors. That is, for any π in S_k,

$$U_\pi \bigotimes_{j=1}^{k} |\psi_j\rangle = \bigotimes_{j=1}^{k} |\psi_{\pi^{-1}(j)}\rangle.$$

We denote the projector onto the completely symmetric subspace of $\mathcal{H}^{\otimes k}$ by $P_{\mathrm{sym}}^{(k)}$. It has rank $\binom{D+k-1}{k}$ and can be expressed as an average over the action unitary representation just described (see, for example, [1]),

$$P_{\mathrm{sym}}^{(t)} = \frac{1}{k!}\sum_{\pi\in S_k} U_\pi.$$

Definition 11. *A (weighted) spherical* t*-design is an ensemble* $(p_k, P_k)_{k=1}^n$ *of* 1*-dimensional projectors* P_k *and probabilities* p_k *such that*

$$\sum_k p_k P_k^{\otimes t} = \frac{1}{\binom{D+t-1}{t}} P_{\mathrm{sym}}^{(t)} = \frac{1}{t!\binom{D+t-1}{t}} \sum_{\pi\in S_t} U_\pi.$$

Note that the isotropic POVM is an ∞-design. We call a t-design *proper* if all the probabilities are equal, $p_k = 1/n$. Note that any t-design is automatically also a t'-design for all $t' < t$. In particular, $\sum_k p_k P_k = \frac{1}{D}\mathbb{1}$, so it makes sense to associate a POVM with every t-design of the form

$$(E_k)_{k=1}^n, \quad \text{with } E_k = D p_k P_k,$$

which, as before, we also call a (weighted or proper) t-design.

It turns out that 4-designs already achieve essentially the same worst–case bias as the isotropic POVM (in the sense that the dimensional dependence is the same). This was discovered by Ambainis and Emerson [5], who showed, invoking a beautiful moment inequality by Berger, that if M_4 is a 4-design POVM then

$$\|\rho-\sigma\|_{\mathrm{M}_4} \geq \frac{1}{3}\|\rho-\sigma\|_2 \geq \frac{1}{3\sqrt{D}}\|\rho-\sigma\|_1. \tag{23}$$

We briefly review their argument, including the Berger inequality, as we need to return to this later on in Sect. 4.

Lemma 12 (Berger [12])**.** *For a real random variable* S*,*

$$\mathbb{E}|S| \geq \frac{(\mathbb{E}S^2)^{3/2}}{(\mathbb{E}S^4)^{1/2}}.$$

Proof. That is just Hölder's inequality, which states that for real random variables f and g, and $\frac{1}{p} + \frac{1}{q} = 1$,

$$\mathbb{E}(fg) \leq \left(\mathbb{E}|f|^p\right)^{1/p} \left(\mathbb{E}|g|^q\right)^{1/q}.$$

Here it is applied with $f = |S|^{2/3}$, $g = |S|^{4/3}$ and $p = 3/2$, $q = 3$. □

Proof (of Eq. (23) – see [5]). For traceless ξ, consider the random variable S which takes value $D\,\mathrm{Tr}(\xi P_k)$ with probability p_k. Then clearly $\mathbb{E}|S| = \|\xi\|_{\mathrm{M}_4}$, and Berger's inequality can be used. The moments are easy calculations, using the fact that the POVM is a 4-design. First, the second moment,

$$\begin{aligned}
\mathbb{E}S^2 &= \sum_k p_k D^2 \left(\mathrm{Tr}(\xi P_k)\right)^2 \\
&= \sum_k p_k D^2 \,\mathrm{Tr}\left((\xi \otimes \xi)(P_k \otimes P_k)\right) \\
&= D^2 \,\mathrm{Tr}\left((\xi \otimes \xi)\frac{2}{D(D+1)} P^{(2)}_{\mathrm{sym}}\right) \\
&= \frac{D^2}{D(D+1)} \,\mathrm{Tr}\left((\xi \otimes \xi)(\mathbb{1} + F)\right) = \frac{D}{D+1}\,\mathrm{Tr}(\xi^2),
\end{aligned}$$

where F is the swap operator, that is $F = U_s$, where s is the non–identity element of S_2, and we have made use of $\mathrm{Tr}(\xi) = 0$. Similarly,

$$\begin{aligned}
\mathbb{E}S^4 &= \sum_k p_k D^4 \left(\mathrm{Tr}(\xi P_k)\right)^4 \\
&= \sum_k p_k D^4 \,\mathrm{Tr}\left((\xi \otimes \xi \otimes \xi \otimes \xi)(P_k \otimes P_k \otimes P_k \otimes P_k)\right) \\
&= D^4 \,\mathrm{Tr}\left(\xi^{\otimes 4} \frac{24}{D(D+1)(D+2)(D+3)} P^{(4)}_{\mathrm{sym}}\right) \\
&= \frac{D^4}{D(D+1)(D+2)(D+3)} \sum_{\pi \in S_4} \mathrm{Tr}\left(\xi^{\otimes 4} U_\pi\right) \\
&= \frac{D^3}{(D+1)(D+2)(D+3)} \left(6(\mathrm{Tr}(\xi^2))^2 + 3\,\mathrm{Tr}(\xi^4)\right) \leq \left(\frac{D}{D+1}\right)^3 9(\mathrm{Tr}(\xi^2))^2.
\end{aligned}$$

The equality in the last line comes from the fact that there are 3 elements of the permutation group S_4 with a 4–cycle, each giving rise to a term equal to $\mathrm{Tr}(\xi^4)$, and likewise, 6 elements which have 2 2–cycles, each yielding a term equal to $(\mathrm{Tr}(\xi^2))^2$. All other elements of the group have at least one fixed point so the corresponding terms contain a factor of $\mathrm{Tr}(\xi)$, which is zero. The final inequality is just an application of the Cauchy–Schwartz inequality. Thus,

$$\|\xi\|_{\mathrm{M}_4} = \mathbb{E}|S| \geq \frac{1}{3}\sqrt{\mathrm{Tr}(\xi^2)} = \frac{1}{3}\|\xi\|_2 \geq \frac{1}{3\sqrt{D}}\|\xi\|_1.$$

In other words: $\lambda(\mathrm{M}_4) \geq 1/(3\sqrt{D})$. □

It is not known how to construct spherical 4-designs efficiently in general though there must exist a weighted 4-design of cardinality at most $\binom{D+3}{4}^2$. To see this, note that the normalised projector $P_{\mathrm{sym}}^{(4)}/(\mathrm{Tr}\, P_{\mathrm{sym}}^{(4)})$ lies in the convex hull of normalised symmetric product states, which are a subset of the $\binom{D+3}{4}^2 - 1$ dimensional real subspace of trace–one hermitian operators on the symmetric subspace. Carathéodory's theorem [3] tells us that any point in the convex hull of a subset S of a n dimensional space can be written as a convex combination of $n+1$ points from S. Constructions are known for a real vector space of small dimensions [21]. Ambainis and Emerson [5] construct approximate 4-designs which perform almost as well as Eq. (23).

C. Performance of 2-designs. Unfortunately, we have as yet been unable to give the bias for 3-design POVMs, but here we show how to bound it for 2-designs. Consider first a proper 2-design with associated POVM $(E_k = \frac{D}{n} P_k)_{k=1}^n$. I.e.,

$$\frac{1}{n}\sum_k P_k \otimes P_k = \frac{1}{D(D+1)}(\mathbb{1} + F) = \frac{2}{D(D+1)} P_{\mathrm{sym}}^{(2)},$$

with the projector $P_{\mathrm{sym}}^{(2)}$ onto the symmetric subspace of $\mathbb{C}^D \otimes \mathbb{C}^D$ and the swap operator F. Such POVMs are always informationally complete – this will also follow from the theorem below.

An example of a 2-design is a complete set of $D+1$ mutually unbiased bases, which are known to exist if the dimension D is a prime power [10,32]). Let

$$\left\{ (|\psi_s^b\rangle)_{s=1\ldots D} : b = 0, \ldots, D \right\},$$

be the basis vectors of the $D+1$ mutually unbiased bases, where $|\psi_s^b\rangle$ is the s^{th} basis vector of the b^{th} basis. Then the set of basis state projectors $P_s^b = |\psi_s^b\rangle\langle\psi_s^b|$ forms a proper spherical 2-design [25]. It is conjectured that in all dimensions there exist spherical 2-designs with the minimum number $n = D^2$ of elements [28], giving rise to so-called *symmetric informationally complete* (SIC) POVMs. These are only known to exist up to dimension $D = 45$ [28] by numerical results, and for even fewer dimensions up to $D = 19$ by mathematical construction. Zauner's conjecture states that in every dimension there exists a SIC-POVM of a particularly beautiful group symmetric form [33]. We refer to [7,15] for more information.

Let M_2 be any 2-design POVM. Our objective is to prove the relation.

Theorem 13. *For any traceless Hermitian operator ξ,*

$$\|\xi\|_{\mathrm{M}_2} \geq \frac{1}{2}\frac{1}{D+1}\|\xi\|_1. \tag{24}$$

In other words, for any proper 2-design POVM as above, $\lambda(\mathrm{M}_2) \geq \frac{1}{2}\frac{1}{D+1}$.

Proof. Since this is a homogeneous relation, we may w.l.o.g. assume that $\|\xi\|_1 = 2$, meaning that we can write $\xi = \rho - \sigma$ with two orthogonal density operators ρ and σ. Thus, what we need to show is $\|\nu_{\mathrm{M}_2}[\rho] - \nu_{\mathrm{M}_2}[\sigma]\| \geq \frac{1}{D+1}$.

For this, we use Proposition 21 in Appendix A, Ineq. (A1), for the vectors $\vec{p}$ and $\vec{q}$ defined as

$$p_k = \mathrm{Tr}(\rho E_k) = \frac{D}{n}\,\mathrm{Tr}(\rho P_k), \quad q_k = \mathrm{Tr}(\sigma E_k) = \frac{D}{n}\,\mathrm{Tr}(\sigma P_k).$$

Namely,

$$\begin{aligned}\|\nu_{\mathrm{M}_2}[\rho] - \nu_{\mathrm{M}_2}[\sigma]\| &= \|\vec{p} - \vec{q}\|_1 \\ &\geq 1 - n\sum_k \frac{D^2}{n^2}(\mathrm{Tr}(\rho P_k))(\mathrm{Tr}(\sigma P_k)) \\ &= 1 - D^2\frac{1}{n}\sum_k \mathrm{Tr}\left((\rho\otimes\sigma)(P_k\otimes P_k)\right).\end{aligned}$$

Now, the last sum can be evaluated as follows, using the property of spherical 2-design:

$$\begin{aligned}\frac{1}{n}\sum_k \mathrm{Tr}(\rho P_k \sigma P_k) &= \frac{1}{n}\sum_k \mathrm{Tr}\left((P_k\otimes P_k)(\rho\otimes\sigma)\right) \\ &= \frac{1}{D(D+1)}\,\mathrm{Tr}\left((\mathbb{1}+F)(\rho\otimes\sigma)\right) \\ &= \frac{1}{D(D+1)}(\mathrm{Tr}(\rho)\,\mathrm{Tr}(\sigma) + \mathrm{Tr}(\rho\sigma)) = \frac{1}{D(D+1)}.\end{aligned}$$

Inserting this above, we conclude

$$\|\nu_{\mathrm{M}_2}[\rho] - \nu_{\mathrm{M}_2}[\sigma]\|_1 \geq 1 - D^2\frac{1}{D(D+1)} = \frac{1}{D+1},$$

as advertised. □

Theorem 14. *For a POVM* M_2 *which is a weighted* 2*-design the conclusion of Theorem 13 still holds:* $\lambda(\mathrm{M}_2) \geq \frac{1}{2}\frac{1}{D+1}$.

Proof. The idea is to break down the probabilities p_k into smaller but approximately equal values. This increases the number of outcomes of the POVM, but makes it be approximated better and better by a proper 2-design, to which we can apply Theorem 13.

In detail, assume that our weighted 2-design is discrete, with n elements; choose an integer $N \gg 1$, and for each k let $N_k = \lfloor Np_k \rfloor$ and $\epsilon_k = Np_k - N_k$. Define a new weighted 2-design with the same projectors $P_{k\ell} = P_k$ and "uniformised" weights

$$\beta_{k\ell} = \begin{cases} \epsilon_k/N & \text{for } \ell = 0, \\ 1/N & \text{for } \ell = 1, \ldots, N_k. \end{cases}$$

Then, applying the same proof as in Theorem 13 to this refined 2-design (which has $N+n$ outcomes), we get

$$\begin{aligned}\|\nu_{\mathrm{M}}[\rho]-\nu_{\mathrm{M}}[\sigma]\|_1 &= \|\vec{p}-\vec{q}\|_1 \\ &\geq 1-(N+n)\sum_{k\ell}\beta_{k\ell}^2 D^2(\mathrm{Tr}(\rho P_k))(\mathrm{Tr}(\sigma P_k)) \\ &\geq 1-D^2\frac{N+n}{N}\sum_{k\ell}\beta_{k\ell}\,\mathrm{Tr}\left((\rho\otimes\sigma)(P_k\otimes P_k)\right) \\ &= 1-\frac{D^2}{D(D+1)}\frac{N+n}{N}\,\mathrm{Tr}\left[(\mathbb{1}+F)(\rho\otimes\sigma)\right] \\ &= 1-\frac{D}{D+1}\left(1+\frac{n}{N}\right)\to\frac{1}{D+1},\end{aligned}$$

where we have used $\beta_{k\ell}\leq 1/N$ in the third line. □

Note that the factor of $1/(D+1)$ in the bound (24) is essentially best possible (up to a constant independent of D), as the example of $D+1$ mutually unbiased bases shows. Indeed, if the two states ρ and σ are distinct elements of one of the bases, then the measured output distributions for all the D other bases are the same, namely uniform, while in their proper basis the trace distance remains 2, so $\|\nu_{\mathrm{M}}[\rho]-\nu_{\mathrm{M}}[\sigma]\|_1=\frac{2}{D+1}$, and hence $\lambda(\mathcal{M})\leq\frac{1}{D+1}$.

Similarly, for a SIC-POVM with D^2 operators $\left(\frac{1}{D}P_k\right)$ it is easily verified that two states from the POVM, i.e. for instance $\rho=P_1$ and $\sigma=P_2$, have trace norm difference $\|\rho-\sigma\|_1=\frac{2D}{D+1}$, while $\|\nu_{\mathrm{M}}[\rho]-\nu_{\mathrm{M}}[\sigma]\|_1=\frac{2}{D+1}$, so $\lambda(\mathcal{M})\leq\frac{1}{D}$.

4. Local POVMs

Consider now a multipartite system $\mathcal{H}=\mathcal{H}_1\otimes\mathcal{H}_2\otimes\cdots\otimes\mathcal{H}_n$, of local Hilbert spaces $\mathcal{H}_j$ of dimension d_j. (The total space's dimension is denoted $D=d_1d_2\cdots d_n$ in this section.) This partition suggests various classes of POVMs due to restrictions of locality. For instance, let **LO** be the class of all *local operations*, i.e. tensor product measurements:

$$\mathbf{LO}=\left\{\left(E_{k_1}^{(1)}\otimes\cdots\otimes E_{k_n}^{(n)}\right):(E_{k_j}^{(j)})\text{ POVM on }\mathcal{H}_j\right\}.$$

More generally, **LOCC** is the class of measurements that can be implemented by local operations and classical communication between the parties. **SEP** are the separable POVMs, i.e.

$$\mathbf{SEP}=\left\{\left(E_k^{(1)}\otimes\cdots\otimes E_k^{(n)}\right):E_k^{(j)}\geq 0,\ \sum_k E_k^{(1)}\otimes\cdots\otimes E_k^{(n)}=\mathbb{1}\right\}.$$

Finally, there is the class of **PPT** POVMs: denoting the transpose operation (with respect to any basis) by T, it is

$$\mathbf{PPT}=\left\{(E_k)\ \text{POVM}:\ \forall k\forall I\subset[n]\ \left(\bigotimes_{i\in I}T\otimes\bigotimes_{i\notin I}\mathrm{id}\right)E_k\geq 0\right\},$$

i.e. all POVM elements have to be PPT with respect to every bipartition of the n-party system.

It is not hard to see that

$$\mathbf{LO} \subset \mathbf{LOCC} \subset \mathbf{SEP} \subset \mathbf{PPT},$$

and all inclusions are known to be strict, at least if the dimension is large enough (see [18] and [19]). The corresponding symmetric convex bodies of operators are denoted

$$\mathbb{LO} \subset \mathbb{LOCC} \subset \mathbb{SEP} \subset \mathbb{PPT}.$$

These are interesting examples of POVM classes since we know due to so-called quantum data hiding [14,26,31] that $\|\xi\|_{\mathbb{M}}$ for them can be much smaller than $\|\cdot\|_1$. Indeed, it was shown in these references that in a bipartite system $\mathbb{C}^d \otimes \mathbb{C}^d$, the states $\sigma = \frac{\mathbb{1}+F}{d(d+1)}$ and $\alpha = \frac{\mathbb{1}-F}{d(d-1)}$, i.e. the (normalised) projectors onto the symmetric and antisymmetric subspace, respectively, obey

$$\left\| \frac{1}{2}\rho - \frac{1}{2}\sigma \right\|_{\mathbb{PPT}} = \frac{2}{d+1}.$$

(In [14] more general statements of this type for n-partite systems can be found.) Consequently, $\lambda(\mathbb{PPT}) \leq \frac{2}{d+1}$. The next result shows that this bound is quite sharp:

Lemma 15. *For any operator ξ on an n-partite system,*

$$\|\xi\|_{\mathbf{SEP}} \geq \frac{2}{2^{n/2}}\|\xi\|_2 \geq \frac{2}{2^{n/2}}\frac{1}{\sqrt{D}}\|\xi\|_1.$$

In particular, $\lambda(\mathbb{SEP}) \geq \frac{2}{2^{n/2}}\frac{1}{\sqrt{D}}$; for a bipartite system, we find $\lambda(\mathbb{SEP}) \geq \frac{1}{\sqrt{D}}$.

Proof. Gurvits and Barnum [20] have shown that for a bipartite system, within the set of Hermitian operators, the unit ball of the Hilbert-Schmidt norm centred on the identity operator contains only separable operators. More generally they proved in an n-partite system, that the ball of radius $2^{1-n/2}$ around the identity is fully separable [20].

It follows immediately that all the POVMs in the set $\left\{ (E, \mathbb{1} - E) : \|2E - \mathbb{1}\|_2 \leq 2^{1-n/2} \right\}$ are separable. It is easy to see that the corresponding symmetric convex body (see Lemma 2) is the ball of radius $2^{1-n/2}$ in the Hilbert-Schmidt norm around the origin and so this is a subset of $\mathbb{SEP}$.

From this inclusion, and the fact that the Hilbert–Schmidt norm is self–dual,

$$\|\xi\|_{\mathbf{SEP}} = \max_{E \in \mathbb{SEP}} \mathrm{Tr}\,(E\xi) \geq \max_{\|E\|_2 \leq 2^{1-n/2}} \mathrm{Tr}\,(E\xi) = \frac{2}{2^{n/2}}\|\xi\|_2,$$

concluding the proof, if we recall $\|\xi\|_1 \leq \sqrt{D}\|\xi\|_2$. □

We now come to the main technical result of the present section, showing that this order of magnitude goes through all the way to $\mathbb{LO}$, indeed, a particular tensor product POVM on a bipartite system is already almost as good as the class of all separable POVMs, in terms of the constant of domination. Note that Proposition 8 gives us the local POVM with the largest λ: namely, by symmetrising over all unitaries $U = U_A \otimes U_B$, drawn from the product of the local Haar measures, we find that for any tensor product POVM $\mathrm{M_A} \otimes \mathrm{M_B}$, we have $\lambda\,(\mathrm{M_{UU}}) \geq \lambda\,(\mathrm{M_A} \otimes \mathrm{M_B})$, where $\mathrm{M_{UU}}$ denotes the tensor product of the isotropic POVMs on the two subsystems.

Theorem 16. *For any two states ρ and σ on a bipartite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, let $\xi = \rho - \sigma$. Then,*

$$\|\xi\|_{\mathrm{M_{UU}}} \geq \frac{1}{\sqrt{153}}\|\xi\|_2 \geq \frac{1}{\sqrt{153D}}\|\xi\|_1,$$

where $D = d_A d_B$ is the Hilbert space dimension. Consequently, $\lambda\,(\mathrm{M_{UU}}) \geq 1/\sqrt{153D}$.

Proof. We do exactly the same as in Subsect. 3B, only that we have now a POVM on $\mathcal{H}_A \otimes \mathcal{H}_B$ of the form

$$(D\mathrm{d}\varphi\mathrm{d}\psi\,|\varphi\rangle\langle\varphi| \otimes |\psi\rangle\langle\psi|),$$

so S is the variable

$$S = D\,\mathrm{Tr}((|\varphi\rangle\langle\varphi| \otimes |\psi\rangle\langle\psi|)\xi),$$

and the bias of the estimation based on the outcome is $\mathbb{E}|S|$, as before in Subsect. 3 B.

We use Berger's inequality, Lemma 12 again, for which we need the second and fourth moment. Because now we randomise independently over $\mathcal{H}_A$ and $\mathcal{H}_B$, we get

$$\begin{aligned}
\mathbb{E}S^2 &= \frac{2^2 d_A^2 d_B^2}{d_A(d_B+1)d_B(d_B+1)}\,\mathrm{Tr}\left((\Pi_{\mathrm{sym}}^{AA} \otimes \Pi_{\mathrm{sym}}^{BB})(\xi^{AB} \otimes \xi^{AB})\right),\\
\mathbb{E}S^4 &= \frac{24^2 d_A^4 d_B^4}{d_A(d_A+1)(d_A+2)(d_A+3)d_B(d_B+1)(d_B+2)(d_B+3)}\\
&\quad \times \mathrm{Tr}\left((\Pi_{\mathrm{sym}}^{AAAA} \otimes \Pi_{\mathrm{sym}}^{BBBB})(\xi^{AB} \otimes \xi^{AB} \otimes \xi^{AB} \otimes \xi^{AB})\right),
\end{aligned}$$

where the superscripts remind one of the systems these operators act on.

Expanding the projectors into the permutations of two, respectively four, elements, we get

$$\mathbb{E}S^2 = \frac{d_A d_B}{(d_A+1)(d_B+1)}\left(\mathrm{Tr}(\xi_A^2) + \mathrm{Tr}(\xi_B^2) + \mathrm{Tr}(\xi^2)\right), \tag{25}$$

where $\xi_A = \mathrm{Tr}_B(\xi)$ and $\xi_B = \mathrm{Tr}_A(\xi)$, because we get terms with $\mathbb{1}^{AA} \otimes \mathbb{1}^{BB}$, $\mathbb{1}^{AA} \otimes F^{BB}$, $F^{AA} \otimes \mathbb{1}^{BB}$ and $F^{AA} \otimes F^{BB}$.

The fourth moment is considerably more complex: looking at

$$\begin{aligned}
\mathbb{E}S^4 &= \frac{d_A^3 d_B^3}{(d_A+1)(d_A+2)(d_A+3)(d_B+1)(d_B+2)(d_B+3)}\\
&\quad \times \sum_{\pi,\sigma\in S_4} \mathrm{Tr}\left((U_\pi^{AAAA} \otimes U_\sigma^{BBBB})\xi^{\otimes 4}\right),
\end{aligned} \tag{26}$$

we see that we need to calculate – or at least reasonably upper bound – the trace terms $\mathrm{Tr}\left((U_\pi^{AAAA} \otimes U_\sigma^{BBBB})\xi^{\otimes 4}\right)$. In Appendix C, Lemma 26 we show that

$$\begin{aligned}
&\sum_{\pi,\sigma\in S_4} \mathrm{Tr}\left((U_\pi^{AAAA} \otimes U_\sigma^{BBBB})\xi^{\otimes 4}\right)\\
&\leq 153(\mathrm{Tr}(\xi^2))^2 + 126(\mathrm{Tr}(\xi^2))(\mathrm{Tr}(\xi_A^2)) + 126(\mathrm{Tr}(\xi^2))(\mathrm{Tr}(\xi_B^2))\\
&\quad + 9(\mathrm{Tr}(\xi_A^2))^2 + 9(\mathrm{Tr}(\xi_B^2))^2 + 30(\mathrm{Tr}(\xi_A^2))(\mathrm{Tr}(\xi_B^2))\\
&\leq 153\left(\mathrm{Tr}(\xi^2) + \mathrm{Tr}(\xi_A^2) + \mathrm{Tr}(\xi_B^2)\right)^2.
\end{aligned}$$

Plugging this into Eq. (26), we find

$$\mathbb{E}S^4 \leq \left(\frac{d_A d_B}{(d_A+1)(d_B+1)}\right)^3 153\left(\mathrm{Tr}(\xi^2)+\mathrm{Tr}(\xi_A^2)+\mathrm{Tr}(\xi_B^2)\right)^2. \tag{27}$$

Now we conclude as in the single-system case: by virtue of Eqs. (25) and (27),

$$\begin{aligned}\|\xi\|_{\mathrm{M_{UU}}} &= \mathbb{E}|S| \\ &\geq \sqrt{\frac{(\mathbb{E}S^2)^3}{\mathbb{E}S^4}} \\ &\geq \frac{1}{\sqrt{153}}\sqrt{\mathrm{Tr}(\xi^2)+\mathrm{Tr}(\xi_A^2)+\mathrm{Tr}(\xi_B^2)} \\ &\geq \frac{1}{\sqrt{153}}\|\xi\|_2 \geq \frac{1}{\sqrt{153D}}\|\xi\|_1,\end{aligned}$$

and we are done. □

Remark 17. From the proof we see that, just as in the single-system case of Subsect. 3 B, it is enough for the local measurements to be 4-designs.

Corollary 18. *The constants of domination, for locality-restricted measurements on a $d \times d$-system, are in the following relations:*

$$\frac{1}{\sqrt{153}d} \leq \lambda\left(\mathrm{M_{UU}}\right) \leq \lambda(\mathbb{LO}) \leq \lambda(\mathbb{LOCC}) \leq \lambda(\mathbb{SEP}) \leq \lambda(\mathbb{PPT}) \leq \frac{2}{d+1}. \tag{28}$$

For separable measurements we have the even tighter bounds,

$$\frac{1}{d} \leq \lambda(\mathbb{SEP}) \leq \lambda(\mathbb{PPT}) \leq \frac{2}{d+1}. \tag{29}$$

Proof. The first inequality in (28) is just Theorem 16, the chain is by inclusion of the sets of POVMs, with the last bound following from the data hiding states α_d and σ_d, the (appropriately normalised) projections onto the (anti-)symmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$ – see [31,14] and [26]. By Lemma 15 finally, $\lambda(\mathbb{SEP}) \geq \frac{1}{\sqrt{D}} = \frac{1}{d}$. □

Remark 19. The first inequality (28) in Corollary 18 proves a conjecture about the optimal bias achievable with LOCC measurements ([26, Conjecture 7]. Compare also with [31], where a bias of order $1/d^2$ was proven using a particular informationally complete measurement, and it was suggested there that better POVMs might exist.

This result shows that in a very strong sense the original data hiding states, the symmetric and anti-symmetric subspace projections, are essentially optimal: up to a constant factor they achieve the best available bias, which is $\Theta(1/d)$.

Remark 20. The ℓ^2-bound in Theorem 16 has another notable consequence for data hiding: observing that for orthogonal states ρ and σ,

$$\|\rho-\sigma\|_2 = \sqrt{\mathrm{Tr}(\rho^2)+\mathrm{Tr}(\sigma^2)} \geq \max\left\{\|\rho\|_2, \|\sigma\|_2\right\},$$

we conclude that data hiding states *have* to be highly mixed. If one of them has rank bounded by r, say, Theorem 16 places a lower bound of $1/13r$ on the bias achievable by LOCC measurements.

Indeed, all known constructions of data hiding states endow them with considerable entropy (comparable to or larger than the size of the "shares"), see [14,22,31]. Our bound tells us that this has to be so to guarantee security of the scheme. We intend to return to this issue on a separate occasion.

5. Certainty Relations

The results on $\lambda(\mathrm{M_U})$ for the isotropic POVM, tensor products of isotropic POVMs, and 2-designs have nice interpretations as "certainty relations" in the sense of Sanchez-Ruiz [29]. Namely, for a complete set of $D+1$ mutually unbiased bases in $\mathbb{C}^D$ with associated basis measurements B_k, he shows that for any pure state $\varphi = |\varphi\rangle\langle\varphi|$,

$$(D+1)\log\frac{D+1}{2} \leq \sum_{k=0}^{D} S_2\left(\nu_{\mathrm{B}_k}[\varphi]\right) \leq (D+1)\log D - \log(D-1), \tag{30}$$

where $S_2\left(\nu_{\mathrm{B}_k}[\varphi]\right) = -\log\sum_x |\langle x|\varphi\rangle|^4$ is the Rényi entropy of order 2 for the orthonormal basis $\{|1\rangle, \ldots |D\rangle\}$. The right hand side of Eq. (30) is referred to as a certainty relation, and intuitively states that for the chosen measurements there exists no pure state that will lead to maximum entropy for all measurements simultaneously. It quantifies the fact (quite natural, after a moment of thought) that not all the tomographic data from measuring those bases is equally informative in the sense of Shannon. The certainty relation of [29] also holds for the Shannon entropy. Let M be the POVM formed by measuring in one of the $D+1$ bases at random. Using the concavity of the log, the certainty relation can then be rewritten as

$$\log\left(D(D+1)\right) - S_2\left(\nu_{\mathrm{M}}[\varphi]\right) \geq \frac{1}{D+1}\log(D-1).$$

From our results in the previous section, we can infer similar certainty relations. First of all, from Theorem 13 we get the following more general but weaker bound for any proper 2-design POVM with n outcomes:

$$\begin{aligned}
\log n - S_2\left(\nu_{\mathrm{M}}[\varphi]\right) &\geq \log n - S\left(\nu_{\mathrm{M}}[\varphi]\right) \\
&= D\left(\nu_{\mathrm{M}}[\varphi]\|\nu_{\mathrm{M}}[\mathbb{1}/D]\right) \\
&\geq \frac{1}{2\ln 2}\|\nu_{\mathrm{M}}[\varphi - \mathbb{1}/D]\|^2 \\
&\geq \frac{1}{4\ln 2}\frac{d-1}{D(D+1)^2} \geq \frac{1}{6\ln 2}\frac{1}{(D+1)^2},
\end{aligned}$$

where $D(\cdot\|\cdot)$ is the classical relative entropy and the second inequality follows from the Pinsker inequality $D(\mu\|\nu) \geq \frac{1}{2\ln 2}\|\mu-\nu\|^2$ (see [4], for example, for definitions of the relative entropy between measures).

For uni- and bipartite 4-designs, in particular the isotropic POVMs, we get considerably better bounds, due to the appearance of the Hilbert-Schmidt norm. Consider any ensemble of quantum states, $\rho = \sum_x p_x\rho_x$. For the Shannon mutual information between the preparation variable X (distributed according to p_x) and the measurement

outcome given by $\mathcal{U}$,

$$\begin{aligned} I(X:\mathrm{M_U}) &= \sum_x p_x D\left(\nu_{\mathrm{M_U}}[\rho_x] \| \nu_{\mathrm{M_U}}[\rho]\right) \\ &\geq \sum_x p_x \frac{1}{2\ln 2} \left\| \nu_{\mathrm{M_U}}[\rho_x] - \nu_{\mathrm{M_U}}[\rho] \right\|^2 \\ &\geq \sum_x p_x \frac{1}{18\ln 2} \|\rho_x - \rho\|_2^2 \\ &= \frac{1}{18\ln 2}\left(\sum_x p_x \operatorname{Tr}(\rho_x^2) - \operatorname{Tr}(\rho^2)\right) \\ &= \frac{1}{18\ln 2}\left(S_L(\rho) - \sum_x p_x S_L(\rho_x)\right). \end{aligned} \tag{31}$$

In other words, we get a lower bound on the accessible information of the ensemble in terms of so-called "linear entropies" $S_L(\rho) = 1 - \operatorname{Tr}(\rho^2)$. In the above derivation we have used the well-known relation between mutual information and relative entropy, the Pinsker inequality and Eq. (23).

A particularly interesting case is that of a pure state ensemble $\rho_x = |\varphi_x\rangle\langle\varphi_x|$: all the $S_L(\rho_x)$ are zero, so we get a positive lower bound for the accessible information

$$I_{\mathrm{acc}}(\{p_x, \varphi_x\}) \geq I(X:\mathrm{M_U}) \geq \frac{1}{18\ln 2}\left(1 - \operatorname{Tr}(\rho^2)\right),$$

which is a small but positive constant, depending only on ρ. It turns out that the best possible lower bound on the accessible information in terms solely of ρ is known: it is the so-called *subentropy* $Q(\rho)$ of Jozsa, Robb and Wootters [24], attained on a particular ensemble decomposition of ρ, named after Ebenezer Scrooge. Incidentally, for this ensemble *all* complete (i.e., rank-1) POVMs have the same information gain. It is largest on the maximally mixed state, and bounded by $\frac{1-\gamma}{\ln 2} \approx .6099$, where γ is Euler's constant [24].

For bipartite systems we furthermore obtain a lower bound for $I_{\mathrm{acc}}^{\mathrm{LOCC}}(\cdot)$, that is the accessible information when we are restricted to performing LOCC measurements. This bound is obtained by using Theorem 16 to lower bound $I(X:\mathrm{M_{UU}})$ – the mutual information when the locally unitarily invariant continuous POVM is used. This quantity is studied as a lower bound on the locally accessible information in [30] (where it is denoted $\Lambda_L(\{p_x, \varphi_x\})$). Unlike the subentropy, this quantity depends on the ensemble (rather than the ensemble average alone) even when it is a pure state ensemble. However, in [30] it is interpreted differently as the average of the mutual information over all complete product basis measurements. Since some measurements of this form cannot be performed by LOCC, the authors (unnecessarily) restrict their claim that it is a lower bound on the locally accessible information to bipartite systems of $2 \times n$ dimensions (where it is known that any complete product basis measurement can be performed by LOCC). This is unnecessary because, as described in Sect. 4, $I(X:\mathrm{M_{UU}})$ is also the mutual information yielded by the protocol where Alice and Bob independently measure according to the unitarily invariant continuous POVM and share their results (which is clearly accomplished by LOCC). As noted in [30], this bound is saturated by Scrooge ensembles.

No general closed form is known for $I(X : \mathrm{M}_{\mathrm{UU}})$ (although some special cases are derived in [30]) so it is useful to note that by using the same derivation as in (31), but invoking Theorem 16, we get that for an arbitrary ensemble on a bipartite system,

$$I_{\mathrm{acc}}^{\mathrm{LOCC}}(\{p_x, \rho_x\}) \geq I(X : \mathrm{M}_{\mathrm{UU}}) \geq \frac{1}{306 \ln 2}\left(S_L(\rho) - \sum_x p_x S_L(\rho_x)\right). \tag{32}$$

It is worth noting that in the case of an ensemble of pure states this lower bound, unlike $I(X : \mathrm{M}_{\mathrm{UU}})$, depends only on the ensemble average. Hence we get a lower bound of

$$Q^{\mathrm{LOCC}}(\rho) := \inf_{\rho=\sum_x p_x \varphi_x} I_{\mathrm{acc}}^{\mathrm{LOCC}}(\{p_x, \varphi_x\}) \geq \frac{1}{306 \ln 2}\left(1 - \mathrm{Tr}(\rho^2)\right)$$

on the LOCC-subentropy of ρ.

6. Conclusion

We have introduced a formalism of norms on states/density operators linked to their (pairwise) distinguishability by a given, restricted, class of measurements. This allows us to study the relation between these norms in convex geometric terms. We went on to investigate the constants of domination for the resulting norms with respect to the well-known trace norm: for a single measurement we looked at the isotropic POVM, 4- and 2-designs. Furthermore, we considered several classes of locally restricted measurements, such as LOCC or PPT POVMs. The results here have strong connection to data hiding: indeed, we proved that up to a constant factor the hiding states of [31] achieve already the best possible bias. We leave many questions open, such as the eventual determination of the locally accessible information and better bounds on the constants of domination. More importantly, one ought to be able to obtain more information on the geometry of the convex bodies $\mathbb{M}$ and the unit balls of $\|\cdot\|_{\mathbb{M}}$ – here we only compared them with the trace and the Hilbert-Schmidt norms, but it would be interesting to get more insight into their geometric shape. It is an intriguing open question regarding single measurements where to place 3-design POVMs relative to 2- and 4-designs.

Acknowledgements. AW thanks the members of the Pavia Quantum Information group for an enjoyable afternoon in October 2007, where he had occasion to discuss some of the questions of the present paper, when they were still in a nascent state. In particular the feedback of G. M. D'Ariano, G. Chiribella and M. F. Sacchi, and their suggestions regarding the use of symmetry, are gratefully acknowledged. Ashley Montanaro provided the pointer to the paper by Ambainis and Emerson, and provided the example mentioned in Appendix A. WM would like to thank Dan Shepherd for a useful discussion about groups and diagrams.

WM was supported by the U.K. EPSRC. SW was supported by NSF grant number PHY-04056720. AW was supported by the U.K. EPSRC through the "QIP IRC" and an Advanced Fellowship, by a Royal Society Wolfson Merit Award and by the European Commission through IP "QAP". The Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence programme.

Appendix A: An ℓ_1-Inequality for Probability Vectors and Density Operators

Proposition 21. *For probability vectors $\vec{p}$, $\vec{q}$ in $\mathbb{R}^n$ (i.e. $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$, and likewise for q_i),*

$$\|\vec{p} - \vec{q}\|_1 \geq 1 - n\, \vec{p} \cdot \vec{q}, \tag{A1}$$

where on the left is the statistical distance between the distributions, namely the ℓ_1-norm of their difference, and on the right we have the usual Euclidean inner product of vectors.

Corollary 22 (Quantum case)**.** *Ineq. (A1) has a straightforward quantum generalisation: for any two density operators ρ and σ on an n-dimensional Hilbert space,*

$$\|\rho - \sigma\|_1 \geq 1 - n \operatorname{Tr}(\rho\sigma), \tag{A2}$$

where now on the left is the trace norm, and on the right is the Hilbert-Schmidt inner product on operator space.

This actually follows from the classical case, as follows: ρ is diagonalised in some basis, with a probability vector $\vec{p}$ along the diagonal. Denote the dephasing operation in this basis by $\mathcal{E}$ – it is a CPTP map with $\mathcal{E}(\rho) = \rho$. Denoting $\sigma' = \mathcal{E}(\sigma)$, which is now diagonalised in the same basis, with a probability vector $\vec{q}$ along the diagonal, we now have

$$\frac{1}{2}\|\rho - \sigma\|_1 \geq \frac{1}{2}\|\rho - \sigma'\|_1 \quad \text{and} \quad \operatorname{Tr}(\rho\sigma) = \operatorname{Tr}(\rho\sigma'),$$

so all we need to prove is

$$\frac{1}{2}\|\rho - \sigma'\|_1 \geq 1 - n \operatorname{Tr}(\rho\sigma').$$

But because of

$$\frac{1}{2}\|\rho - \sigma'\|_1 = \frac{1}{2}\|\vec{p} - \vec{q}\|_1 \quad \text{and} \quad \operatorname{Tr}(\rho\sigma') = \vec{p} \cdot \vec{q},$$

this is precisely (A1). □

Proof of Proposition 21. We use the well-known relation between trace distance and fidelity [16]:

$$\frac{1}{2}\|\vec{p} - \vec{q}\|_1 \geq 1 - \sum_i \sqrt{p_i q_i},$$

hence we are done once we show

$$2\left(1 - \sum_i \sqrt{p_i q_i}\right) \geq 1 - n \sum_i p_i q_i,$$

which – introducing the shorthand $t_i = \sqrt{p_i q_i}$ – is equivalent to

$$\sum_i t_i \leq \frac{1}{2} + \frac{1}{2} n \sum_i t_i^2.$$

Now, for fixed $s = \sum_i t_i \leq 1$, the right hand side here is minimal for $t_1 = \ldots = t_n = \frac{s}{n}$, in which case it reduces to $\frac{1}{2} + \frac{1}{2}s^2$, which is indeed always $\geq s$. □

Remark 23. Ineq. (A1) becomes false when introducing a factor $c < 1$ on the left hand side. for sufficiently large n. Ashley Montanaro [personal communication] pointed out to us the following class of examples:

Consider $\vec{p} = \left(x, 0, \frac{1-x}{n-2}, \ldots, \frac{1-x}{n-2}\right)$ and $\vec{q} = \left(0, x, \frac{1-x}{n-2}, \ldots, \frac{1-x}{n-2}\right)$, which have $c\|\vec{p} - \vec{q}\|_1 = 2cx$, whereas $1 - n\vec{p} \cdot \vec{q} = 1 - \frac{n}{n-2}(1-x)^2 \sim 2x + x^2$ for large n.

Appendix B: An Integral Over the Unit Sphere

Lemma 24. *Let P and Q be mutually orthogonal projectors of rank a and b, respectively, in $\mathbb{C}^d$. Then, for the uniform distribution on the unit vectors $|\psi\rangle = \sum_{j=1}^d \psi_j |j\rangle \in \mathbb{C}^d$,*

$$\begin{aligned}\mathbb{E}\left|\frac{1}{2a}\operatorname{Tr}(\psi P) - \frac{1}{2b}\operatorname{Tr}(\psi Q)\right| &= d\int \mathrm{d}\psi \left|\frac{1}{2a}\sum_{j=1}^{a}|\psi_j|^2 - \frac{1}{2b}\sum_{j=a+1}^{a+b}|\psi_j|^2\right| \\ &= 1 - \frac{1}{a+b}\sum_{\substack{k=0,\ldots a-1\\ \ell=0,\ldots b-1}} p^k(1-p)^\ell\binom{k+\ell}{k},\end{aligned}$$

where $p = a/(a+b)$.

Proof. Introduce a random Gaussian vector $|\varphi\rangle \sim \mathcal{N}_{\mathbb{C}^d}(0,1)$ [11], i.e. $|\varphi\rangle = \frac{1}{\sqrt{2d}}\sum_{j=1}^d (\alpha_j + i\beta_j)|j\rangle$ with independent Gaussian distributed real and imaginary parts $\alpha_j, \beta_j \sim \mathcal{N}(0,1)$ of zero mean and unit variance. In particular, $\mathbb{E}\langle\varphi|\varphi\rangle = 1$.

Now, using this and the unitary invariance of the distribution of $|\varphi\rangle$, we see

$$\begin{aligned}\mathbb{E}\left|\frac{1}{2a}\operatorname{Tr}(\psi P) - \frac{1}{2b}\operatorname{Tr}(\psi Q)\right| &= \mathbb{E}_\varphi\left(\langle\varphi|\varphi\rangle\mathbb{E}_\psi\left|\frac{1}{2a}\operatorname{Tr}(\psi P) - \frac{1}{2b}\operatorname{Tr}(\psi Q)\right|\right) \\ &= \mathbb{E}_\varphi\left|\frac{1}{2a}\operatorname{Tr}(\varphi P) - \frac{1}{2b}\operatorname{Tr}(\varphi Q)\right| \\ &= \frac{1}{2}\mathbb{E}_{\alpha_j,\beta_j\sim\mathcal{N}(0,1)}\left|\frac{1}{2a}\sum_{j=1}^{a}(\alpha_j^2+\beta_j^2)\right. \\ &\qquad \left. -\frac{1}{2b}\sum_{j=a+1}^{a+b}(\alpha_j^2+\beta_j^2)\right| \\ &= \frac{1}{2}\mathbb{E}_{X,Y}\left|\frac{1}{2a}X - \frac{1}{2b}Y\right|.\end{aligned}$$

The sums of squares of Gaussian components occurring here are well-studied, and known under the name of χ^2-distributions:

$$\sum_{j=1}^{a}(\alpha_j^2+\beta_j^2) =: X \sim \chi^2_{2a}, \quad \sum_{j=a+1}^{a+b}(\alpha_j^2+\beta_j^2) =: Y \sim \chi^2_{2b},$$

their probability density being given by

$$\begin{aligned}\Pr\{X \in [x; x+\mathrm{d}x]\} &= \frac{1}{2(a-1)!}(x/2)^{a-1}e^{-x/2}\mathrm{d}x, \\ \Pr\{Y \in [y; y+\mathrm{d}y]\} &= \frac{1}{2(b-1)!}(y/2)^{b-1}e^{-y/2}\mathrm{d}y.\end{aligned}$$

This allows us to evaluate the latter expectation as follows, denoting the indicator function of a set $\{\ldots\}$ as $\mathbf{1}\{\ldots\}$:

$$\begin{aligned}\frac{1}{2}\mathbb{E}_{X,Y}\left|\frac{1}{2a}X-\frac{1}{2b}Y\right| &= \frac{1}{2}\mathbb{E}_{X,Y}\left(\int \mathrm{d}r\,\mathbf{1}\{X/2a\le r\le Y/2b\}\right.\\ &\qquad\left.+\int \mathrm{d}r\,\mathbf{1}\{Y/2b\le r\le X/2a\}\right)\\ &= \frac{1}{2}\int_0^\infty \mathrm{d}r\,(\mathbb{E}\,\mathbf{1}\{X\le 2ar, Y\ge 2br\}+\mathbb{E}\,\mathbf{1}\{X\ge 2ar, Y\le 2b\})\\ &= \frac{1}{2}\int_0^\infty \mathrm{d}r\,(\Pr\{X\le 2ar\}\Pr\{Y\ge 2br\}\\ &\qquad+\Pr\{X\ge 2ar\}\Pr\{Y\le 2br\})\\ &= \frac{1}{2}\int_0^\infty \mathrm{d}r\,\Pr\{X\ge 2ar\}+\frac{1}{2}\int_0^\infty \mathrm{d}r\,\Pr\{Y\ge 2br\}\\ &\quad-\int_0^\infty \mathrm{d}r\,\Pr\{X\ge 2ar\}\Pr\{Y\ge 2br\}.\end{aligned}$$

Using the χ^2 densities, the probabilities under the integrals are easily evaluated:

$$\Pr\{X\ge 2ar\}=e^{-ar}\sum_{k=0}^{a-1}\frac{(ar)^k}{k!},\quad \Pr\{Y\ge 2br\}=e^{-br}\sum_{\ell=0}^{b-1}\frac{(br)^\ell}{\ell!}.$$

This finally gives

$$\begin{aligned}\mathbb{E}\left|\frac{1}{2a}\operatorname{Tr}(\psi P)-\frac{1}{2b}\operatorname{Tr}(\psi Q)\right| &= \frac{1}{2}\mathbb{E}_{X,Y}\left|\frac{1}{2a}X-\frac{1}{2b}Y\right|\\ &= \frac{1}{2}+\frac{1}{2}-\int_0^\infty \mathrm{d}r\,e^{-r(a+b)}\sum_{\substack{k=0,\ldots a-1\\ \ell=0,\ldots b-1}}\frac{(ar)^k(br)^\ell}{k!\ell!}\\ &= 1-\frac{1}{a+b}\sum_{\substack{k=0,\ldots a-1\\ \ell=0,\ldots b-1}}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^\ell\binom{k+\ell}{k},\end{aligned}$$

where we have used the integral for the Gamma function. □

We will also need the following small lemma:

Lemma 25. *Let S_k denote $\sum_{l=0}^k 2^{-(k+l)}\binom{k+l}{l}$. We claim that for integers $k\ge 0$, $S_k=1$.*

Proof. Using the well known 'addition formula' $\binom{n}{m}=\binom{n-1}{m}+\binom{n-1}{m-1}$,

$$S_{k+1}=\sum_{l=0}^{k+1}2^{-(1+k+l)}\binom{k+l}{l}+\sum_{l=0}^{k+1}2^{-(1+k+l)}\binom{k+l}{l-1} \tag{B1}$$

$$=\frac{1}{2}S_k+2^{-(2k+2)}\binom{2k+1}{k+1}+\sum_{l=0}^{k}2^{-(2+k+l)}\binom{k+l+1}{l} \tag{B2}$$

$$=\frac{1}{2}S_k+2^{-(2k+2)}\binom{2k+1}{k+1}+\frac{1}{2}S_{k+1}-\frac{1}{2}2^{-(2k+2)}\binom{2k+2}{k+1} \tag{B3}$$

so

$$S_{k+1} = S_k + 2^{-(2k+2)} \left(2\binom{2k+1}{k+1} - \binom{2k+2}{k+1} \right) = S_k,$$

where the final equality is due to the addition formula and the symmetry $\binom{2k+1}{k+1} = \binom{2k+1}{k-1}$. To complete the proof we note that $S_0 = 1$. □

Appendix C: Upper Bounds on Certain Traces

Lemma 26. *Let ξ be a traceless Hermitian operator on a bipartite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. Let $P^{(4)}_{\mathrm{sym}\,A}$ and $P^{(4)}_{\mathrm{sym}\,B}$ denote the projector onto the completely symmetric subspace of $\mathcal{H}_A^{\otimes 4}$ and $\mathcal{H}_B^{\otimes 4}$, respectively.*

Then, with the shorthands $t := \mathrm{Tr}(\xi^2)$, $a := \mathrm{Tr}(\xi_A^2)$ and $b := \mathrm{Tr}(\xi_B^2)$, where $\xi_A = \mathrm{Tr}_B(\xi)$ and $\xi_B = \mathrm{Tr}_A(\xi)$,

$$\mathrm{Tr}\left(\left(P^{(4)}_{\mathrm{sym}\,A} \otimes P^{(4)}_{\mathrm{sym}\,B} \right) \xi^{\otimes 4} \right) \leq \frac{1}{4!^2} \left(153t^2 + 126ta + 126tb + 9a^2 + 9b^2 + 30ab \right). \tag{C1}$$

The proof is conceptually simple but a little long. We write the projection operators as averages over the unitary operators which permute the four subsystems. Defining, for permutations $\pi \in S_4$, the representation

$$U^A_\pi := \sum_{\mathbf{j} \in \{1,\ldots,d\}^m} \bigotimes_{i=1}^{4} |j_i\rangle^A_{\pi(i)} \langle j_i|^A_i,$$

where $\{|j\rangle^A_i\}_{1 \leq j \leq d}$ is an orthonormal basis for the i^{th} copy of $\mathcal{H}_A$ in $\mathcal{H}_A^{\otimes 4}$, and defining U^B_π similarly:

$$\mathrm{Tr}\left(\left(P^{(4)}_{\mathrm{sym}\,A} \otimes P^{(4)}_{\mathrm{sym}\,B} \right) \xi^{\otimes 4} \right) = \frac{1}{24!^2} \sum_{\pi \in S_4, \sigma \in S_4} \mathrm{Tr}\left(U^A_\pi \otimes U^B_\sigma \xi^{\otimes 4} \right).$$

Clearly $(\pi, \sigma) \to U^A_\pi \otimes U^B_\sigma$ is a representation of $S_4 \times S_4$. $S_4 \times S_4$ has a subgroup consisting of all the elements of the form (g, g), which we'll denote by R.

If $(\pi', \sigma') = r^{-1}(\pi, \sigma) r$ for some $r \in R$, we write $(\pi', \sigma') \overset{R}{\sim} (\pi, \sigma)$ and note that the corresponding terms are equal since

$$\begin{aligned} \mathrm{Tr}\left(U^A_{\pi'} \otimes U^B_{\sigma'} \xi^{\otimes 4} \right) &= \mathrm{Tr}\left((U^A_\pi \otimes U^B_\sigma)(U^A_g \otimes U^B_g) \xi^{\otimes 4} (U^A_{g^{-1}} \otimes U^B_{g^{-1}}) \right) \\ &= \mathrm{Tr}\left(U^A_\pi \otimes U^B_\sigma \xi^{\otimes 4} \right). \end{aligned}$$

Essentially, conjugation by an element of R corresponds to a permutation of the identical ξ operators, and therefore leaves the term unchanged.

The set of all $24!^2$ terms is partitioned by the equivalence relation $\overset{R}{\sim}$ with the terms in each subset all equal to each other. We shall refer to these subsets as the R-conjugacy classes of $S_4 \times S_4$. Clearly, the R-conjugacy classes form a finer partition of $S_4 \times S_4$ than the normal conjugacy classes.

By demonstrating an appropriate upper-bound for the terms in each R-conjugacy class, and calculating the size of each class, we will prove the upper bound (C1).

Tensor Diagrams. Let us establish an orthonormal basis $\{|i\rangle_A\}$ ($\{|i\rangle_B\}$) for $\mathcal{H}_A$ ($\mathcal{H}_B$). In this basis, we can write ξ in component form thus $\xi^{k,l}_{i,j} = \langle k|_A \otimes \langle l|_B \xi |i\rangle_A \otimes |j\rangle_B$.

We would like to demonstrate upper bounds for terms of the form

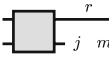
$$\mathrm{Tr}\left(U^A_\pi \otimes U^B_\sigma \xi^{\otimes 4}\right) = \xi^{a\pi(1),b\sigma(1)}_{a1,b1} \xi^{a\pi(2),b\sigma(2)}_{a2,b2} \xi^{a\pi(3),b\sigma(3)}_{a3,b3} \xi^{a\pi(4),b\sigma(4)}_{a4,b4}, \qquad \text{(C2)}$$

where the ai and bi ($i \in \{1, 2, 3, 4\}$) are dummy variables to be contracted over according to the Einstein summation convention. Using indices in our calculations would be rather messy and confusing. Instead we use the ingenious tensor diagrams of Penrose [27]:

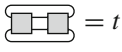
We denote our bipartite Hermitian operator ξ by  . The "terminals" of this diagram correspond to indices like so

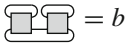
$\xi^{k,l}_{i,j} =$ 

Joining the terminals with "wires" denotes contraction of the corresponding indices:

$\xi^{k,l}_{r,j}\xi^{r,m}_{p,q} =$ 

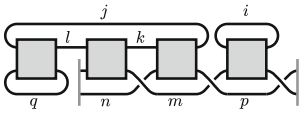
$\xi_A := \mathrm{Tr}_B(\xi) =$  , $\quad \xi_B := \mathrm{Tr}_A(\xi) =$  ,

$\mathrm{Tr}(\xi) =$  $= 0$, $\quad \mathrm{Tr}(\xi^2) =$  $= t$,

$\mathrm{Tr}(\xi_A^2) =$  $= a$, $\quad \mathrm{Tr}(\xi_B^2) =$  $= b$.

In an effort to keep the diagrams tidy and compact, we sometimes use a pair of vertical grey lines, one with wires entering from the right and the other with a matching set of wires entering from the left. A diagram with this feature is to be read as equivalent to the diagram one obtains by identifying the grey lines parallel to join the matching wires. It should not be confused with the bars drawn *across* wires (by Penrose and others) to denote (anti-)symmetrization.

Here is an example showing how a diagram corresponds to a particular term of the form (C2):



$= \xi^{j,q}_{l,q}\xi^{l,p}_{k,m}\xi^{k,n}_{j,p}\xi^{i,m}_{i,n}$.

In Fig. 1 we provide a table with a diagram representative of each of the R-conjugacy classes organised by the conjugacy class of $S_4 \times S_4$ which contains it.

The size of each R-conjugacy class is written to the right of the corresponding diagram. An upper bound is given and diagrams which are identically 0 (by virtue of having a factor of $\mathrm{Tr}(\xi) = 0$) are drawn in a lighter shade of grey.

Proofs of upper bounds. We give bounds for the terms shown in the upper-right triangle of Fig. 1. Bounds for those terms below the diagonal follow from these by exchanging the roles of the parties. We will make repeated use of the Cauchy-Schwarz inequality for the Hilbert-Schmidt inner product,

Lemma 27. $|\mathrm{Tr}(A^\dagger B)|^2 \leq (\mathrm{Tr}(A^\dagger A))(\mathrm{Tr}(B^\dagger B))$.

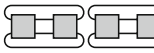
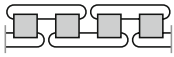
Let P denote a positive semidefinite hermitian operator. We have the inequality $\mathrm{Tr}(P^2) \leq (\mathrm{Tr}(P))^2$ (by the spectral decomposition of P for example). From this fact and the Cauchy-Schwarz inequality it follows that

Lemma 28. *If P and Q are both positive semidefinite, then* $\mathrm{Tr}(PQ) \leq (\mathrm{Tr}(P))(\mathrm{Tr}(Q))$.

Third, since the partial transpose map is selfadjoint,

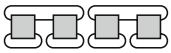
Lemma 29. *The quantities t, a and b are unchanged if we replace ξ with ξ^Γ.*

Proof of of Lemma 26. We go through the types one by one.

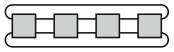
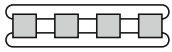
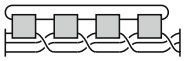
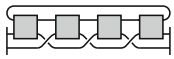
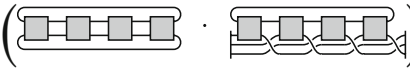
(2,2):(2,2)  $= (\mathrm{Tr}(\xi^2))^2 = t^2$. To show that the same bound applies to  , we note that it can be written as $\mathrm{Tr}\left((\mathrm{Tr}_A(Z))^2\right)$, where

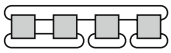
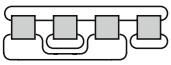
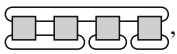
$$Z = (\xi \otimes \mathbb{1}_C)(\mathbb{1}_A \otimes |\Phi\rangle\langle\Phi|)(\xi \otimes \mathbb{1}_C)$$

and $|\Phi\rangle = \sum_{i=1}^d |i\rangle_B \otimes |i\rangle_C$. Since $Z = ((\xi \otimes \mathbb{1}_C)(\mathbb{1}_A \otimes |\Phi\rangle))\,((\xi \otimes \mathbb{1}_C)(\mathbb{1}_A \otimes |\Phi\rangle))^\dagger$, it is positive semidefinite, and as such $\mathrm{Tr}\left((\mathrm{Tr}_A(Z))^2\right) \leq (\mathrm{Tr}(Z))^2$. The result follows by noting that $\mathrm{Tr}(Z) = t$.

(2,2):(1,1,1,1)  $= (\mathrm{Tr}(\xi_A^2))^2 = a^2$.

(2,1,1):(2,1,1)  $= ab$.

(4):(4) Noting that ξ^2 is positive semidefinite, and applying Lemma 28, we get  $= \mathrm{Tr}(\xi^4) \leq (\mathrm{Tr}(\xi^2))^2 = t^2$. The partial-transpose of ξ, ξ^Γ, has the diagrammatic representation  (we choose to take the transpose on Bob's system). Substituting, this for ξ in  results in the diagram  $= (\mathrm{Tr}(\xi^\Gamma))^4$, so Lemma 29 shows that the same bound applies here. The Cauchy-Schwarz inequality yields  $= \mathrm{Tr}\left((\xi^\Gamma)^2(\xi^2)^\Gamma\right) \leq \sqrt{(\mathrm{Tr}(\xi^\Gamma))^4(\mathrm{Tr}\left((\xi^2)^\Gamma\right))^2} =$  , which can be seen to be $\leq t^2$ because of the previous two bounds.

(4):(2,1,1)  $= \mathrm{Tr}\left((\mathrm{Tr}_B(\xi^2))\xi_A^2\right) \leq (\mathrm{Tr}(\xi^2))(\mathrm{Tr}(\xi_A^2)) = ta$, by Lemma 28.  $= \mathrm{Tr}\left(\xi(\xi_A \otimes \mathbb{1}_B)\xi(\xi_A \otimes \mathbb{1}_B)\right) \leq \mathrm{Tr}\left(\xi(\xi_A^2 \otimes \mathbb{1}_B)\xi\right) =$  , where we have used the Cauchy-Schwarz inequality.

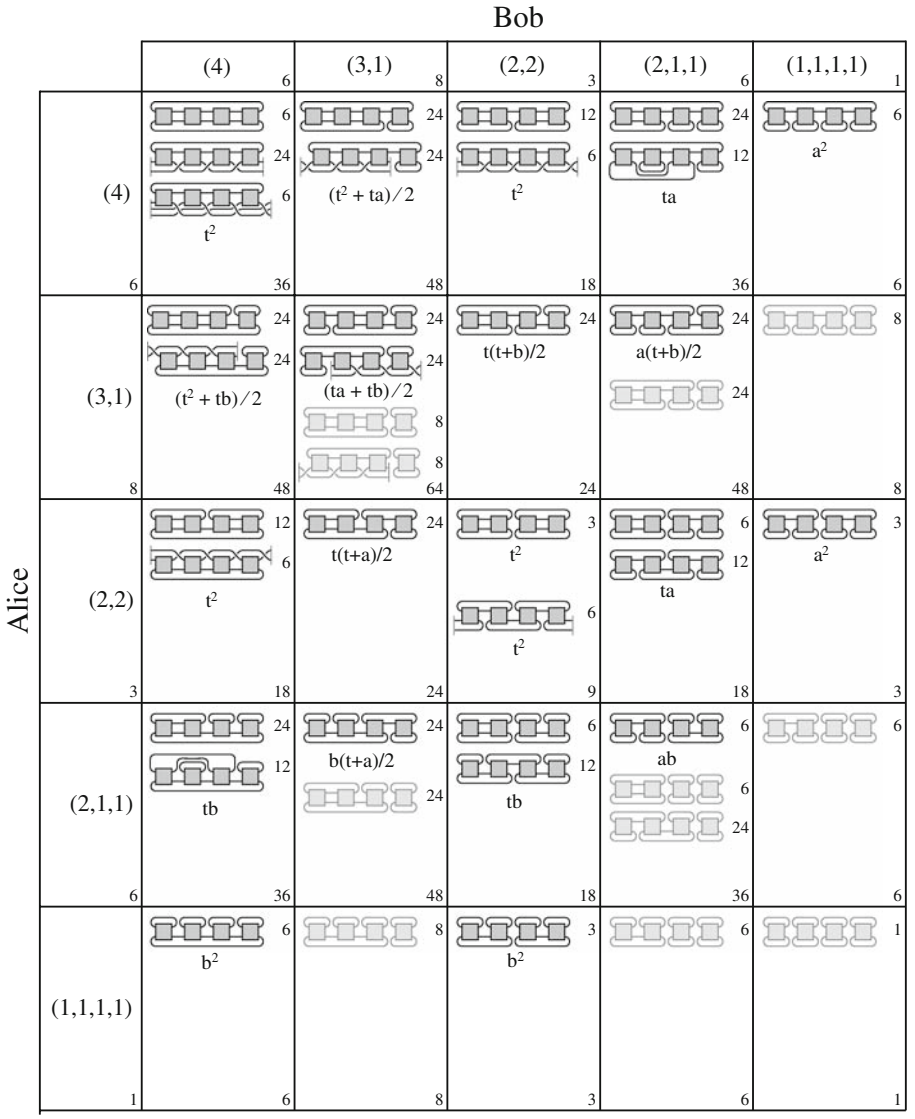
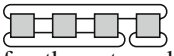
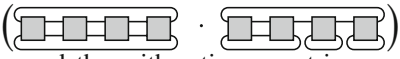
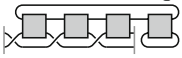
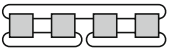
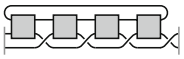
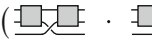
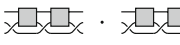
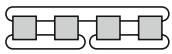


Fig. 1. Sizes and upper-bounding expressions of the R-conjugacy classes. The faded diagrams are identically zero (because they contain a factor of $\mathrm{Tr}(\xi)$)

(4):(3,1)  $\leq$  . Using the results for these two diagrams and the arithmetic-geometric mean inequality we can bound this expression by $t(t+a)/2$, as was claimed.  is given by substituting ξ^{Γ} into the previous diagram, so by Lemma 29 the previous bound applies.

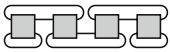
(4):(2,2)  $= \mathrm{Tr}(\mathrm{Tr}_B(\xi^2))^2 \leq t^2$. For the other diagram we use the Cauchy-Schwarz inequality:  $\leq \big(\big(\mathrm{Tr}\,\big($  $\big)\big(\mathrm{Tr}\,\big($  $\big)\big)\big)^{1/2} =$  $\leq t^2$.

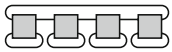
(2,2):(2,1,1)  $= ta$. For the other diagram in this class it is useful to define $Y_B := \mathrm{Tr}_A\,(\xi(\xi_A \otimes \mathbb{1}_B))$. We define Y_A similarly but with the roles of the parties reversed.

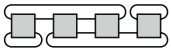
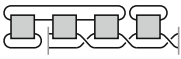
$$\begin{aligned}\mathrm{Tr}(Y_B^2) &= \mathrm{Tr}\,((\mathrm{Tr}_A\,(\xi(\xi_A \otimes \mathbb{1}_B))) \cdot Y_B) \\ &= \mathrm{Tr}\,(\xi(\xi_A \otimes \mathbb{1}_B)(\mathbb{1}_A \otimes Y_B)) \\ &= \mathrm{Tr}\,(\xi(\xi_A \otimes Y_B)) \\ &\leq \sqrt{(\mathrm{Tr}(\xi^2))(\mathrm{Tr}(\xi_A^2))(\mathrm{Tr}(Y_B^2))},\end{aligned}$$

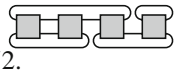
and therefore

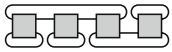
$$\mathrm{Tr}(Y_B^2) \leq (\mathrm{Tr}(\xi^2))(\mathrm{Tr}(\xi_A^2)) = ta.$$

Similarly $\mathrm{Tr}(Y_A^2) \leq tb$. Hence,  $= \mathrm{Tr}(Y_B^2) \leq ta$.

(4):(1,1,1,1)  $= \mathrm{Tr}(\xi_A^4) \leq (\mathrm{Tr}(\xi_A^2))^2 = a^2$.

(3,1):(3,1)  $= \mathrm{Tr}\left((\xi_A \otimes \mathbb{1}_B)\xi^2(\mathbb{1}_A \otimes \xi_B)\right) = \mathrm{Tr}\left(\xi^2(\xi_A \otimes \xi_B)\right)$. Using the Cauchy-Schwarz inequality we upper bound this by $\sqrt{(\mathrm{Tr}(\xi^4))(\mathrm{Tr}(\xi_A^2))(\mathrm{Tr}(\xi_B^2))}$, which in turn is bounded by $(\mathrm{Tr}(\xi^2))\sqrt{(\mathrm{Tr}(\xi_A^2))(\mathrm{Tr}(\xi_B^2))} \leq (ta + tb)/2$, using arithmetic-geometric mean inequality at the end.  is given by substituting ξ^Γ into the previous diagram, so by Lemma 29 the same bound applies.

(3,1):(2,2)  $= \mathrm{Tr}\left((\mathrm{Tr}_B(\xi^2))Y_A\right) \leq \sqrt{(\mathrm{Tr}\left(\mathrm{Tr}_B(\xi^2)\right))^2\,(\mathrm{Tr}(Y_A^2))} \leq t(t+b)/2$.

(3,1):(2,1,1)  $= \mathrm{Tr}(\xi_A^2 Y_A) \leq \sqrt{(\mathrm{Tr}(\xi_A^4))(\mathrm{Tr}(Y_A^2))} \leq a(t+b)/2$.

Now, collecting terms according to the multiplicities found in the table of Fig. 1, we conclude the proof. □

Remark 30. Note that for every pair of conjugacy classes of permutations, all the types falling into the corresponding box in Fig. 1 share the same upper bound.

References

1. Fulton, W., Harris, J.: *Representation Theory: A First Course*. Berlin Heidelberg NewYork: Springer, 1991
2. Dudley, R.M.: *Real Analysis and Probability*. Cambridge: Cambridge University Press, 2002
3. Rockafellar, R.T.: *Convex Analysis*. Princeton, NJ: Princeton University Press, 1997
4. Cover, T., Thomas, J.: *Elements of Information Theory* (Second Edition). NewYork: John Wiley and Sons (2006)
5. Ambainis, A., Emerson, J.: *Quantum t-designs: t-wise independence in the quantum world*. In: Proc. 22nd Annual IEEE Conference on Computational Complexity (CCC'07), 129–140, available at http://arxiv.org/abs/quant-ph:0701126v2, 2007

6. Appleby, D.M., Dang, H.B., Fuchs, C.A.: *Physical Significance of Symmetric Informationally-Complete Sets of Quantum States*. http://arxiv.org/abs/:0707.2071v1[quant-ph], 2007
7. Appleby, D.M.: SIC-POVMs and the Extended Clifford Group. In J. Math. Phys. **46**, 052107 (2005)
8. Audenaert, K.M.R., Calsamiglia, J., Munoz-Tapia, R., Bagan, E., Masanes, Ll., Acín, A., Verstraete, F.: Discriminating States: The Quantum Chernoff Bound. Phys. Rev. Lett. **98**, 160501 (2007); Nussbaum, M., Szkoła, A.: The Chernoff lower bound for symmetric quantum hypothesis testing. Ann. Stat. **37**, no. 2, 1040–1057 (2009)
9. Ballester, M.A., Wehner, S., Winter, A.: State Discrimination with Post-Measurement Information. In: IEEE Trans. Inf. Theory **54**, no. 9, 4183–4198 (2008)
10. Bandyopadhyay, S., Boykin, P.O., Roychowdhuri, V., Vatan, F.: A New Proof for the Existence of Mutually Unbiased Bases. Algorithmica **34**, 512–528 (2002)
11. Bennett, C.H., Hayden, P., Leung, D., Shor, P.W., Winter, A.: Remote preparation of quantum states. IEEE Trans. Inf. Theory **51**(1), 56–74 (2005)
12. Berger, B.: The Fourth Moment Method. SIAM J. Comput. **24**(6), 1188–1207 (1997)
13. Bolthausen, E.: An Estimate of the Remainder in a Combinatorial Central Limit Theorem. Zeits. für Wahrsch. Und Verw. Geb. **66**, 387–405 (1984)
14. Eggeling, T., Werner, R.F.: Hiding Classical Data in Multipartite Quantum States. Phys. Rev. Lett. **89**(9), 097905 (2002)
15. Flammia, S.: On SIC-POVMs in Prime Dimensions. J. Phys. A. Math. Gen. **39**, 10901–10907 (2006)
16. Fuchs, C.A., van de Graaf, J.: Cryptographic distinguishability measures for quantum-mechanical states. IEEE Trans. Inf. Theory **45**(4), 1216–1227 (1999)
17. Grassl, M.: On SIC-POVMs and MUBs in Dimension 6. J. Phys. A: Math. Gen. **39**, 13483–13493 (2006)
18. Bennett, C., DiVincenzo, D., Fuchs, C., Mor, T., Rains, E., Shor, P., Smolin, J., Wootters, W.: Quantum nonlocality without entanglement. Phys. Rev. A **59**(2), 1070–1091 (1999)
19. Horodecki, M., Horodecki, P., Horodecki, R.: Separability of Mixed States: Necessary and Sufficient Conditions. Phys. Lett. A **223**, 1–8 (1996)
20. Gurvits, L., Barnum, H.: Largest separable balls around the maximally mixed bipartite quantum state. Phys. Rev. A **66**, 062311 (2002); Gurvits, L., Barnum, H.: Separable balls around the maximally mixed multipartite quantum states. Phys. Rev. A **68**, 042312 (2003)
21. Hardin, R.H., Sloane, N.J.A.: McLaren's Improved Snub Cube and Other New Spherical Designs in Three Dimensions. Discrete and Comput. Geom. **15**, 429–441 (1996)
22. Hayden, P., Leung, D., Shor, P.W., Winter, A.: Randomizing Quantum States: Constructions and Applications. Commun. Math. Phys. **250**(2), 371–391 (2004)
23. Helstrom, C.W.: *Quantum Detection and Estimation Theory*. New York: Academic Press, (1976); Holevo, A.S.: Statistical decision theory for quantum systems. J. Multivariate Anal. **3**(4), 337–394, (1973)
24. Jozsa, R., Robb, D., Wootters, W.K.: Lower bound for accessible information in quantum mechanics. Phys. Rev. A **49**(2), 668–677 (1994)
25. Klappenecker, A., Roetteler, M.: Mutually Unbiased Bases are complex spherical 2-designs. In: Proc. ISIT 2005, Piscataway, NJ: IEEE, 2005, pp. 1740–1744
26. Matthews, W., Winter, A.: On the Chernoff distance for asymptotic LOCC discrimination of bipartite quantum states. Commun. Math. Phys. **285**(1), 161–174 (2009); http://arxiv.org/abs/0710.4113v2[quant-ph], 2008
27. Penrose, R., Rindler, W.: *Spinors and Space-Time, Vol. 1: Two-spinor calculus and relativistic fields*. Cambridge: Cambridge University Press, 1986
28. Renes, J.M., Blume-Kohout, R., Scott, A.J., Caves, C.M.: Symmetric Informationally Complete Quantum Measurements. J. Math. Phys. **45**, 2171 (2004)
29. Sanchez-Ruiz, J.: Entropic uncertainty and certainty relations for complementary observables. Phys. Lett. A **173**(3), 233–239 (1993); Improved bounds in the entropic uncertainty and certainty relations for complementary observables. Phys. Lett. A **201**(2–3), 125–131 (1995)
30. Sen De, A., Sen, U., Lewenstein, M.: Distillation protocols that involve local distinguishing: Composing upper and lower bounds on locally accessible information. Phys. Rev. A **74**, 052332 (2006)
31. Terhal, B.M., DiVincenzo, D.P., Leung, D.: Hiding Bits in Bell States. Phys. Rev. Lett. **86**(25), 5807–5810 (2001); DiVincenzo, D.P., Leung, D., Terhal, B.M.: Quantum data hiding. IEEE Trans. Inf. Theory **48**(3), 580–599 (2002)
32. Wootters, W.K., Fields, B.D.: Optimal state-determination by mutually unbiased measurements. Ann. Phys. **191**, 363–381 (1989)
33. Zauner, G.: *Quantum Designs – Foundations of a Non-Commutative Theory of Designs* (in German). Ph.D. thesis, Universität Wien (1999)

Communicated by M.B. Ruskai