IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Quantum state determination: estimates for information gain and some exact calculations

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2007 J. Phys. A: Math. Theor. 40 10887

(http://iopscience.iop.org/1751-8121/40/35/011)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 171.66.16.144
The article was downloaded on 03/06/2010 at 06:12

Please note that terms and conditions apply.

IOP Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **40** (2007) 10887–10902 doi:10.1088/1751-8113/40/35/011

Quantum state determination: estimates for information gain and some exact calculations

M K Patra

Department of Computing and Mathematics, University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 1797, Australia

E-mail: manas@scm.uws.edu.au

Received 14 December 2006, in final form 12 July 2007
Published 14 August 2007
Online at stacks.iop.org/JPhysA/40/10887

Abstract
It is known that mutually unbiased bases (MUBs), whenever they exist, are optimal in an information theoretic sense for the determination of the unknown state of a quantum ensemble. Such bases may not exist for most dimensions. The present paper deals with information gain in some generalizations and approximations of MUBs. We give estimates of the information loss (relative to MUBs) in these suboptimal choice of bases. For some generalization of MUBs we give exact calculations for the information gain. It is calculated directly in terms of transition probabilities among the measurement bases. We also give the formal solutions for the problem of quantum state tomography in these cases[1].

PACS numbers: 03.65.Wj, 03.67.−a, 02.10.Yn, 02.50.−r

1. Introduction

The state of a quantum system is completely specified by a ray in a complex Hilbert space $\mathcal{H}$ (infinite dimensional, in general) or, more generally, by a density matrix. A density matrix is a positive operator on $\mathcal{H}$ with unit trace. Thus, a density matrix has nonnegative eigenvalues whose sum equals 1. In particular, it is Hermitian. In this work, we restrict ourselves to finite dimensional spaces.

Often the state of the quantum system is not known *a priori* and has to be determined by certain tests. The problem of state tomography is to determine the state from the outcomes of measurements carried out on an *ensemble* of identically prepared systems. A general quantum measurement scheme is given by a set of positive operator valued measures satisfying certain constraints [3]. Similarly, in the context of tomographic quantum cryptography [1] we have

[1] Some of the results in this paper was reported in the proceedings of EQIS 2005. A large part of this work was done at CQCT, Macquarie University, Sydney.

1751-8113/07/3510887+16$30.00 © 2007 IOP Publishing Ltd Printed in the UK

to choose a set of positive operator valued measures (POVM) for measurement. What is the optimal choice of such POVMs? This has been partially answered for projection valued measures (PVM) by an information theoretic analysis [2]. Let us distinguish two problems concerned with general measurements. The first is the problem of estimation [3]: given a measurement procedure and the data find the best estimates of the state that produced the data, assuming some prior distribution on the state. A recent analysis of this aspect may be found in [4]. The second problem which is the primary concern for the present paper may be termed a *design* problem. Given a class of measurements to determine the parameters characterizing the state, find the 'best measurement'. We will see that in this optimization problem the determinant of a certain matrix plays a crucial role. This matrix is of independent interest since its inverse gives the formal solution to the state determination problem. A substantial part of this paper is denoted to the study of this matrix. A brief outline of the paper follows.

The first three sections are more or less reviews of known results from quantum theory and statistical information theory which helps us fix the framework and the notation. In section 2 we state the formal problem of determining the quantum state of an ensemble. In section 3 we review the information content of a measurement. We use the phrase 'information content' (instead of information gain) because the information depends on the measurement and the prior distribution of states and could be negative. However, the *average information* is always non-negative and henceforth by information gain we will imply average information gain.

Section 4 deals with the problem of quantum state determination by projective measurements. In particular, we give formal solutions. The information gain of a class of quantum measurements under the assumption that the prior distributions of the states is *uniform* is proportional to the determinant of a certain matrix Γ [2]. There are some valid objections to this assumption and we discuss its justification. We show that the information gain is maximum for mutually unbiased bases. We also obtain some estimates on the information gain in the general case based on certain determinant approximations [10]. In any case, the matrix Γ is of independent interest as its inverse gives the formal solution to the state reconstruction problem. In section 5 we apply the results of the previous sections to some generalizations of mutually unbiased bases (MUBs). We define a class of bases which appear to be new. We give exact calculations of the determinant of Γ in these cases by investigating its spectrum. We also show the form of its eigenvectors. In particular, we get a concrete algorithm for reconstructing the state from the probabilities. We first apply the general techniques and formulae to some construction of bases in [12]. Then we discuss examples of a new class of bases in small dimensions. The final section discusses some issues potential problems the author aims to investigate in the future.

2. State determination

In this section we look into the problem of quantum state determination. Let $\mathcal{H}$ be a Hilbert space of dimension n. Let $V(\mathcal{H})$ be the set of Hermitian operators on $\mathcal{H}$. The dimension of $V(\mathcal{H})$ as a *real* vector space is n^2. Let $\Omega(\mathcal{H}) \subset V(\mathcal{H})$ be the set of positive operators with trace 1. It is a convex set. The map $L : V(\mathcal{H}) \to V(\mathcal{H})$ such that $L(T) = T - \frac{\mathrm{Tr}(T)}{n} I$ is linear and the image $l(\mathcal{H})$ is the space of Hermitian operators with trace 0. It has dimension $n^2 - 1$. Here I denotes the identity operator. The affine space $l(\mathcal{H}) + I/n$ consists of all Hermitian operators with trace 1. Therefore, a density matrix is completely specified by $n^2 - 1$ parameters.

The state of a quantum system is not directly measurable. A measurement yields only probabilities or more precisely frequencies. We assume for simplicity that all Hermitian operators are observable. Let $\rho \in \Omega(\mathcal{H})$ be a state and A a Hermitian operator. Let $A = \sum_i c_i |\alpha_i\rangle\langle\alpha_i|$ be the spectral decomposition of A. The space of linear operators on $\mathcal{H}$ becomes a Hilbert space of complex dimension n^2 by defining the inner product

$$\langle B, C\rangle = \mathtt{Tr}(B^\dagger C). \tag{1}$$

The corresponding norm is the Frobenius norm. Its restriction to $V(\mathcal{H})$ makes the latter a *real* inner product space of dimension n^2. The probability of obtaining the ith outcome $p_i = \mathtt{Tr}(|\alpha_i\rangle\langle\alpha_i|\rho) = \langle|\alpha_i\rangle\langle\alpha_i|, \rho\rangle$ may be interpreted as projections of the state ρ onto the corresponding 'coordinate' vector $|\alpha_i\rangle\langle\alpha_i|$. It is more convenient to consider the traceless Hermitian operators $\rho - I/n$ instead of ρ. Now, a measurement on an ensemble in some basis can at best give us an estimate of the n^2 probabilities of the possible outcomes. We may thus characterize a measurement of a nondegenerate observable A by the corresponding orthonormal basis in the spectral decomposition. Thus, the measurement of several observables on the subensembles of the original ensemble is equivalent to giving a set of bases. We shall henceforth simply refer to a choice of bases of measurement as a *measurement scheme*. Let N be the number of samples of the quantum system on which the measurement is made. Of the n numbers $f_i = N_i/N$, the relative frequencies of obtaining result i in a measurement in some basis $\mathcal{B}$ only $n-1$ are independent since $\sum_i f_i = 1$. Let $P_i = |\alpha_i\rangle\langle\alpha_i|, i = 1, \ldots, n$ be the projection operators in $\mathcal{B}$. They satisfy $P_i P_j = \delta_{ij} P_i$ and $\sum_i P_i = I$. As vectors in $V(\mathcal{H})$ they are linearly independent. To avoid confusion we call the projection operators corresponding to some basis in the ambient space $\mathcal{H}$ *projectors* when they are considered as elements of $V(\mathcal{H})$. But there are only n of them that are orthogonal. Further, if we have two bases $\mathcal{B}_i$ and $\mathcal{B}_2$, then at most $2n-1$ projectors from $\mathcal{B}_2$ can be independent of those in $\mathcal{B}_1$ due to the relation $\sum_i P_i = I$. Hence to get the $n^2 - 1$ coordinates of ρ we need at least $n+1$ bases such that they contain a maximally independent set in the following sense. For each projector P_i^j from the jth basis let $T_i^j = P_i^j - I/n$ be the corresponding traceless operators. If the collection $\{T_i^j\}$ contain a maximally independent set (cardinality $= n^2 - 1$) then they span $l(\mathcal{H})$. We call such a set of projectors a complete set of measurement bases (CSMB for short). If the number of bases is $n+1$ and they are complete then we call such a set *independent*. Suppose we have two CSMBs $\mathcal{S}_1$ and $\mathcal{S}_2$. If all other conditions are identical, which one should we pick for determining the unknown state of a quantum ensemble? We may assume ideal conditions—perfect preparation procedures, perfect detectors and measuring devices etc—to compare the two. In [2] Fields and Wootters proved that a set of mutually unbiased bases (MUBs) is an optimal choice when the initial distribution is uniform. Two sets of orthonormal bases $\{|\alpha_i\rangle\}$ and $\{|\beta_j\rangle\}$ are said to be mutually unbiased if $|\langle\alpha_i|\beta_j|\rangle|^2 = 1/n$ for any pair of vectors. They further go on to show that such bases exist whenever the dimension n is a power of some prime, extending the earlier work of Ivanovic [5] who demonstrated the existence of MUBs in prime dimension. These works, however, left open the question of the existence of MUBs for n which divides two or more distinct primes, e.g., six. It is now widely believed that MUBs do not exist in such dimensions. However, we can expect CSMBs which approximate MUBs. Then it is natural to ask: how much do we lose due to the approximations? This question is relevant even in the cases where MUBs are known to exist because in more realistic situations the measurement apparatus will only approximately implement the MUBs. However, to answer such questions we must have an appropriate framework in which these questions may be posed and answered precisely and quantitatively. The natural candidate seems to be information theory.

3. Information content of a measurement

In this section we follow [6] to define the information content of an experiment and apply it to the case of the quantum state determination measurements. First let us look at the general formalism. Let $\mathcal{M}$ be a measurement on some system S, not necessarily quantum. We should use the term 'experiment' rather than measurement since the latter seems to imply a single measurement. Let S be characterized by some parameters denoted by θ which will usually be drawn from some subset $\mathbf{\Theta}$ of $\mathbb{R}^k$, the k-dimensional Euclidean space. Let $p(\theta)$ represent the *a priori* probability distribution of the parameters θ. Corresponding to every value of θ there is a probability measure on $\boldsymbol{X}$—the set of possible measurement data which is again a subset of some Euclidean space. We assume that this measure is given by $p(\boldsymbol{x}|\theta)\,\mathrm{d}\boldsymbol{x}$. Consequently, $\int_B p(\boldsymbol{x}|\theta)\,\mathrm{d}\boldsymbol{x}$ is the conditional probability of getting the outcome $\boldsymbol{x}$ in $B \subset \boldsymbol{X}$ given the state θ. Let $p(\boldsymbol{X}) = \int_{\boldsymbol{X}} p(\boldsymbol{x}|\theta)p(\theta)\,\mathrm{d}\theta$ be the probability density of the random variable $\boldsymbol{x}$. Note that we have used the same symbol p for the probability densities of different random variables. This does not mean that they are the same functions. The notation is more convenient and unambiguous if taken in proper context. Moreover, we do not differentiate between a random variable and its values. In an experiment we are often interested in the *posterior* probability $p(\theta|\boldsymbol{x})$. It is the probability density for θ given the measured values $\boldsymbol{x}$. The information content of the measurement $\mathcal{M}$ is defined as

$$\mathfrak{I}(\mathcal{M}, p(\theta), \boldsymbol{x}) \equiv \int p(\theta|\boldsymbol{x}) \log p(\theta|\boldsymbol{x})\,\mathrm{d}\theta - \int p(\theta) \log p(\theta)\,\mathrm{d}\theta. \tag{2}$$

If $p(\theta|\boldsymbol{x}) = 0$ then the integrand is defined to be zero and the logarithm is taken over an arbitrary but fixed base. The justification for this definition is as follows. Consider the term

$$\mathfrak{I}_0 \equiv -\int p(\theta) \log p(\theta)\,\mathrm{d}\theta.$$

It represents the initial entropy capturing our prior uncertainty about the parameter θ. The quantity

$$\mathfrak{I}_1 = -\int p(\theta|\boldsymbol{x}) \log p(\theta|\boldsymbol{x})\,\mathrm{d}\theta$$

gives us the entropy *after* the measurement which resulted in the reading $\mathbf{x}$ of the relevant parameter. We expect that the measurement decreases the initial uncertainty, which is the purpose of any experiment! Thus $\mathfrak{I}_0 - \mathfrak{I}_1$ gives the information gain in the measurement. It could be negative for a poor design. Then we know less (more uncertainty) after the measurement! An important property required of any measure of information is a kind of additivity property. Suppose it is known to the experimenter that θ is found in $\mathbf{\Theta}' \subset \mathbf{\Theta}$ with probability q. Let $\mathcal{I}_1$ be a measure of information corresponding to this knowledge. In the next stage the experimenter is told the value of θ. Let $\mathcal{I}_2$ and $\mathcal{I}_3$ be the amount of information gained in the next phase when this value is in $\mathbf{\Theta}'$ or its complement, respectively. Then the fundamental additive property [7] required of the information measure is that the total information

$$\mathcal{I} = \mathcal{I}_1 + q\mathcal{I}_2 + (1-q)\mathcal{I}_3. \tag{3}$$

Then it is not difficult to show that the information measure $\mathfrak{I}_0$ is unique up to a constant multiple. We do not discuss these points further but refer the reader to any good source on basic information theory (e.g. [7] and [6]) for a discussion in the context of experiments. The information gain will in general depend upon the experiment and the distribution of the data $\boldsymbol{x}$. Thus we may say that one experiment or measurement is more informative than others. Let us calculate information content for some simple measurements in the quantum domain.

Let the dimension $n = 2$. Suppose we have prior information that the state is a pure state $|0\rangle$ or $|1\rangle$ with probability $1/2$. We may therefore model the parameter space as $\Theta = \{0, 1\}$ with $p(0) = p(1) = 1/2$. Then $\mathfrak{I}_0 = -(1/2\log(1/2) + 1/2\log(1/2)) = 1$. The logarithm is taken to the base 2. Now suppose that we choose to make measurement $\mathcal{M}_1$ in the basis $\{|0\rangle, |1\rangle\}$ which is natural, given the prior information. Then the conditional probabilities may be conveniently written in the matrix form, for $i, j \in \{0, 1\}$,

$$p(i|j) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

It is simply the unit matrix of order 2. Thus, if we get the measurement outcome 0 we are sure that the state of the system was $|0\rangle$ etc. Then it is easy to see that $\mathfrak{I}_1(\mathcal{M}_1, i) = 0$ and hence $\mathfrak{I}(\mathcal{M}_1, i) = \mathfrak{I}_0 - \mathfrak{I}_1(\mathcal{M}_1, i) = 1$. Now suppose we perversely choose the basis $|\overset{+}{-}\rangle = \frac{1}{\sqrt{2}}(|0\rangle \overset{+}{-} |-\rangle)$ for measurement $\mathcal{M}_2$. Then the corresponding conditional probability matrix is

$$p(i|j) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Again it is easy to see that the information gain in this case is $\mathfrak{I}(\mathcal{M}_2, i) = 0$. That is, we get no information from $\mathcal{M}_2$. The same conclusion follows if the initial distribution on the set $\{|0\rangle, |1\rangle\}$ is $(p, 1-p)$ for any $0 \leqslant p \leqslant 1$.

Although the information measure defined above depends on the state and may be negative the *average* information

$$\mathfrak{I}(\mathcal{M}, p(\theta)) \equiv \int \mathfrak{I}(\mathcal{M}, p(\theta), \boldsymbol{x}) p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{4}$$

is independent of the state and is nonnegative [6]. Here,the probability density

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x}|\theta) p(\theta)\, \mathrm{d}\theta \tag{5}$$

is the mean probability distributions averaged over θ. Using Bayes' rule

$$p(\boldsymbol{x}|\theta)p(\theta) = p(\theta|\boldsymbol{x})p(x)$$

it is not difficult to show that

$$\mathfrak{I}(\mathcal{M}, p(\theta)) = \iint p(\theta)p(\boldsymbol{x}|\theta) \log\left(p(\boldsymbol{x}|\theta)\right) \mathrm{d}\boldsymbol{x}\, \mathrm{d}\theta - \int p(\boldsymbol{x}) \log\left(p(\boldsymbol{x})\right) \mathrm{d}\boldsymbol{x}. \tag{6}$$

This is the formula we use to estimate the information gained in quantum measurements.

4. Quantum state tomography and MUBs

Given an n-dimensional quantum ensemble in an unknown state, how do we determine its state? This is the problem of quantum state tomography. The state is not directly observable but we may infer it from the probability distributions observed. As mentioned in the introduction, we need (projective) measurement in $n + 1$ *independent* bases to determine the state completely from the observed probabilities. Actually, what we observe are the frequencies of the possible outcomes. Then it is a problem of estimation theory to draw inferences about the probabilities. Thus, the state tomography problem has broadly two theoretical aspects. The first is a design issue. What is the optimal choice of bases? The second aspect is a problem of decision or estimation theory: for a given measurement what is the best possible estimate of the parameters characterizing the state? In this paper we will be mainly concerned with the first aspect. So let us formulate the problem precisely now.

Suppose we are given an ensemble of quantum systems in some unknown state ρ. By an ensemble we mean an unlimited supply of identically prepared quantum systems. Let $\mathcal{B}^1, \ldots, \mathcal{B}^r$ be bases in $\mathcal{H}$ with

$$\mathcal{B}^k = \left\{\mathbb{P}^k_i \equiv |\alpha^k_i\rangle\langle\alpha^k_i|\right\}_{i=1}^n. \tag{7}$$

A basis in the ambient Hilbert space will always mean an orthonormal basis unless specified otherwise. As vectors in the n^2-dimensional space $V(\mathcal{H})$ at most n^2 of the vectors from the above bases can be linearly independent. Due to the relations $\sum_i \mathbb{P}^k_i = I$ for all k we can only have all the n vectors from exactly one basis in any independent set of projectors. If $\mathcal{B} = \bigcup_k \mathcal{B}^k$ contains a maximal independent set then we may choose the first $n-1$ projectors $\mathbb{P}^k_i, i = 1, \ldots, n-1$ and $\frac{I}{n}$ as a basis for $V(\mathcal{H})$. Then $\{T^k_i = \mathbb{P}^k_i - I/n : i = 1, \ldots, n-1$ and $k = 1, \ldots, r$ span $l(\mathcal{H})$, the space of traceless Hermitian operators. Let $s^{kl}_{ij} = \langle\mathbb{P}^k_i, \mathbb{P}^l_j\rangle = \mathtt{Tr}\left(\mathbb{P}^k_i\mathbb{P}^l_j\right)$. The nonnegative numbers s^{kl}_{ij} are the respective transition probabilities among the vectors in the kth and lth bases. Note that $s^{kk}_{ij} = \delta_{ij}$ since each of the bases is orthonormal. A measurement scheme will be called complete if the probability distributions in the set of bases defining the measurement determine the state uniquely. The number of bases in a complete measurement scheme must be at least $n+1$. Since for a complete measurement $\left\{T^k_i\right\}$ span $l(\mathcal{H})$, for any state ρ we must have

$$\rho - I/n = \sum y^k_i\left(\mathbb{P}^k_i - I/n\right) \equiv \sum y^k_i T^k_i, \tag{8}$$

for some numbers y^k_i. Then,

$$\mathtt{Tr}\big((\rho - I/n)T^l_j\big) = p^l_j - 1/n = \sum y^k_i\langle T^k_i, T^l_j\rangle = \sum_{i,k} t^{kl}_{ij} y^k_i. \tag{9}$$

Here

$$t^{kl}_{ij} \equiv \langle T^k_i, T^l_j\rangle = \mathtt{Tr}\left(T^k_i T^l_j\right) = s^{kl}_{ij} - 1/n \tag{10}$$

and p^k_i is the probability of the ith outcome in the measurement in the kth basis. It is easy to see that if the original bases are mutually unbiased then $\langle T^k_i, T^l_j\rangle = 0$ for $k \neq l$, that is, the operators T^k_i and $T^l_j, k \neq l$ are orthogonal when considered as vectors. Let Γ be the $(n^2-1) \times (n^2-1)$ matrix whose entries are given by t^{kl}_{ij}. It is the Gramm matrix (see [15], p20) of the vectors T^k_i. If the vectors T^k_i are independent the matrix Γ is invertible. If we consider the parallelepiped spanned by the vectors T^k_i then t^{kl}_{ij} are the angles between the sides T^k_i and T^l_j. Note also that the components y^k_i may be used as input parameters characterizing the state (denoted by θ earlier). We will denote these by a vector $\boldsymbol{Y}$. If $\boldsymbol{p}$ denotes a vector of dimension n^2-1 whose components are $p^i_j - 1/n$ above then

$$\boldsymbol{Y} = \Gamma^{-1}\boldsymbol{p}.$$

So in principle we can determine the state if we know the probabilities in $n+1$ independent bases. Even if the measurement scheme were complete but not independent we can solve equation (8) since the rank is maximal. In this case Γ^{-1} will denote the generalized or Moore–Penrose inverse (see [16] and [17], p 421). We summarize the above discussion in the following proposition.

Proposition 1. *Let ρ be the state and $\mathcal{B}^i \equiv \left\{|\alpha^i_j\rangle \big| j = 1, \ldots, n\right\}, i = i, \ldots, r$ be bases in $\mathcal{H}$. Let $P^i_j = |\alpha^i_j\rangle\langle\alpha^i_j|$ be the corresponding projectors and define*

$$T^i_j = P^i_j - \frac{I}{n}, \qquad i = 1, \ldots, r \quad \textit{and} \quad j = 1, \ldots, n-1.$$

Then if the corresponding measurement scheme is complete any state ρ is completely determined by the probabilities $p^i_j = \mathtt{Tr}(\rho P^i_j), i = 1, \ldots, r$ and $j = 1, \ldots, n-1$ and the transition probabilities $s^{ik}_{jl} = \mathtt{Tr}(P^i_j P^k_l)$.

By a measurement we mean a collection of several observations in different bases on subensembles of the original ensemble. We picture a massively parallel setup where we have a several measuring devices $\mathcal{D}_k$ for each basis $\mathcal{B}^k$. The original ensemble is divided into large subensembles and tested by each of these $n+1$ groups of devices. For each $k \leqslant n+1$ we get frequencies m^k_i for the ith outcome, $1 \leqslant i \leqslant n-1$ in the kth device group. The numbers m^k_i constitute the measurement data $\boldsymbol{x}$ in (2) and (6). What is a reasonable probability distribution for the m^k_i?

Before answering this question we have to posit some prior distribution of states. This is actually a subtle point [19–21][2]. The first problem that we encounter is that of probability assignments: given that ensemble could be in two states ρ_1 and ρ_2 with probability q and $1-q$ do we put this information in the distribution $p(\theta)$ or do we take $q\rho_1 + q\rho_2$ as the state with probability 1? Both possibilities will yield the same *posterior* distributions. This point is discussed well in [21] and he proposes that the assignment of probabilities should reflect the nature of our *ignorance*. We further add that this ignorance should be assessed by taking the whole picture into account. Specifically, in the problem of state determination we assume that whenever we use some measurement scheme an ensemble of identically prepared systems is given to us. For example, if the preparation device randomly chooses a qubit in a state with $\sigma_z = 1$ or $\sigma_x = 1$ with equal probability then the appropriate state is $\rho = 1/2(1/2(1+\sigma_z)) + 1/2(1/2(1+\sigma_x))$. But when we use the same measurement scheme with a different preparation device or at some other time we could get a different ensemble. To measure the information gain for the measurement scheme we must consider its potential for use with *any* preparation device and this is what we mean by the phrase 'whole picture'. Now, how do we fix the distribution of probabilities for all such preparation devices. At first glance, it seems reasonable to posit a uniform distribution assuming that all states are equally probable. However, the main problem with the uniform distribution is that it depends upon the parametrization we use to characterize the state [22]. We may choose a prior which is least informative (see [20] and the references therein). In classical Bayesian theory the Fisher information metric provides a unique volume element on a Riemannian manifold of probability distributions. This volume element provides the most noninformative prior. The quantum case is further complicated by the fact that there is a nondenumerable family of such metrics. We may however compare the relative 'noninformativity' of some of the well-known metrics as is done in [20]. It turns out that in this comparison the uniform distribution is a rather 'informative'. However, let us point out some of our reasons for staying with a uniform prior. The uniform prior that Slater [20] compares to other distributions is over a standard parametrization of the Bloch sphere for a two-dimensional system. We take different parameterizations afforded by bases in the space $l(\mathcal{H})$ where $\mathcal{H}$ is of arbitrary dimension. These bases in $l(\mathcal{H})$ are the projectors formed by the corresponding bases in $\mathcal{H}$. The parametrization that we use are the coefficients with respect to some such fixed basis in $l(\mathcal{H})$. If we change to another basis the information gain changes by a constant factor that is independent of the measurement scheme. As we see below, the factor which does depend on the measurement scheme is a certain determinant and we take this as a measure of information gain. The point is, we are not interested in the absolute value of the information gain but only that factor which depends on our choice of measurement. The question of prior has greater importance when dealing with the issue of estimation, especially if we follow a Bayesian approach. Further, as

[2] The author is grateful to a referee for bringing this point to his attention and suggesting some references.

pointed out in the introduction, the above determinant is of independent interest and the main thrust of the paper is estimating or calculating it. Last but not least, we appeal to simplicity for the present choice of prior and hope to tackle more complicated priors in future work.

Now coming back to the distribution for the frequencies m_i^k (this is $p(\boldsymbol{x}|\theta)$ in equation 6) we may appeal to the local limit theorem [8] in probability theory which roughly states that for independent identically distributed random variables the probability distribution of their frequencies tends to the normal distribution in the limit $N \to \infty$, N being the number of trials. We must have some prior distribution for the states. Let V be the volume of the parallelepiped spanned by the vectors T_i^k. Assuming a uniform distribution for the states it can be shown that [2] the average information gain in a quantum measurement for state determination is proportional to $\ln V$ plus an additive constant independent of the measurement scheme. The limiting case may also derived as a consequence of the law of large numbers [8]. We will take $\ln V$ as the measure for information content of a quantum test of an ensemble in CSMB and for a CSMB $\mathcal{C}$ write $\mathcal{I}(\mathcal{C})$ for the information gain and $V(\mathcal{C})$ for the corresponding volume. The fact that the volume of the parallelepiped spanned by the bases gives a measure of information may be heuristically seen as follows. We noted earlier that the probability p_i^k is the projection of the unknown state ρ onto the vector P_i^k. Hence, $p_i^k - 1/n$ is the projection onto the 'coordinate axis' T_i^k of the parallelepiped. The observed frequencies are distributed around these projections. The spread of this distribution gives the uncertainty in the measurement. Each basis defines one such spread. The total uncertainty will be the sum of individual uncertainties in each basis. The overlapping of these spreads will add to the uncertainty due to the duplications. Hence, the larger the volume of the parallelepiped defined by the bases the lesser is the overlapping of spreads and hence a reduction in the uncertainty. As shown below, the parallelepiped formed by the MUBs has the maximum volume. The sufficiency part in the next result was already given in [2] but the present approach is different.

Theorem 1. *Information gain $\mathcal{I}(\mathcal{C})$ is maximum if and only if $\mathcal{C}$ consists of mutually unbiased bases.*

Proof. From the preceding discussion, we have to show that the volume $V(\mathcal{C})$ spanned by the vectors T_i^k is maximal iff T_i^k and T_j^l are orthogonal for $k \neq l$. Note first that $\langle T_i^k, T_j^k\rangle = \delta_{ij} - 1/n$. Consider the $(n^2-1)\times(n^2-1)$ matrix $\Gamma(\mathcal{C}) = \left(t_{ij}^{kl}\right) = \langle T_i^k, T_j^k\rangle$ and assume the ordering defined by the pair $\{k, i\}$. This simply means that the matrix consists of $n+1$ blocks γ^{kl}, each a square matrix of size $(n-1)$ such that $\gamma^{kl}(ij) = t_{ij}^{kl}$:

$$\Gamma(\mathcal{C}) = \begin{pmatrix} \gamma^{11} & \gamma^{12} & \dots & \gamma^{1n+1} \\ \gamma^{21} & \gamma^{22} & \dots & \gamma^{2n+1} \\ \vdots & \vdots & \dots & \vdots \\ & & \dots & \gamma^{n+1,n+1} \end{pmatrix}. \tag{11}$$

If we choose any orthonormal basis for $l(\mathcal{H})$ and express T_i^k in this basis. Let $\mathcal{T}$ be the corresponding real matrix of the coefficients, then it is clear that $\mathcal{T}^t\mathcal{T} = \Gamma(\mathcal{C})$, where $\mathcal{T}^t$ is the transpose of $\mathcal{T}$. It follows that $\det\Gamma(\mathcal{C}) = (V(\mathcal{C}))^2$ and $\Gamma(\mathcal{C})$ is positive definite. Thus maximizing $V(\mathcal{C})$ is equivalent to maximizing $\Gamma(\mathcal{C})$. Below we will focus on the latter. From the generalized Fischer–Hadamard inequality [9] it follows that

$$\det\Gamma(\mathcal{C}) \leqslant \det\gamma^{11}\cdots\det\gamma^{n+1,n+1}. \tag{12}$$

The right-hand side is the product of the determinants of the diagonal blocks in $\Gamma(\mathcal{C})$. Now, if the T_i^k and T_j^l are orthogonal for $k \neq l$ then the off-diagonal blocks are all zero matrices and the equality holds in equation (12). This proves the sufficiency part.

The equality holds in (12) only if the following condition is satisfied [9]. Let R be the $(n+1)\times(n+1)$ matrix such that $R(ij)=1$ if $\gamma^{ij}\neq 0$ and $R(ij)=0$ otherwise. Then the equality holds if and only if there is a permutation matrix of order $n+1$ such that PRP^{-1} is triangular. Since Γ is symmetric and $P^{-1}=P^t$ it follows that if PRP^{-1} is triangular it must be diagonal. The operation $R\to PRP^{-1}$ permutes the diagonal elements of R among themselves. Hence, PRP^{-1} is diagonal iff all off-diagonal elements are zero. That is, $\gamma^{ij}=0$ for $i\neq j$. That is, the original bases are mutually unbiased. The necessity is proved. □

The above theorem gives an upper bound. A natural question is: how tight is the bound? This pertains to the estimation of the information content in bases which are independent but not mutually unbiased. We now give an estimate of the relative loss due to such a non-optimal choice. First let us compute the determinant in the case of a set of $n+1$ MUBs. We have only diagonal blocks in the matrix Γ. Recall that a diagonal block $\gamma^{kk}(ij)=\langle\mathbb{P}^k_i-1/n,\mathbb{P}^k_j-1/n\rangle=\delta_{ij}-1/n$. The following notation will be used in the rest of the paper. Let J_{rs} denote the $r\times s$ matrix all whose entries are 1. Further, we let $J_r=J_{rr}$ and I_r denote the identity matrix of order r. Thus all the diagonal blocks γ^{kk} are equal to $I_{n-1}-J_{n-1}/n$. Sometimes we suppress the subscripts if the dimensions are clear from the context. We also write Γ instead of $\Gamma(\mathcal{C})$ when the set of measurement bases $\mathcal{C}$ is fixed. Let Γ_0 be the submatrix of Γ consisting of the diagonal blocks.

Lemma 1. $\det(\Gamma_0)=\frac{1}{n^{n+1}}$.

Proof. First note that $J^2_{n-1}=(n-1)J_{n-1}$. The eigenvalues of J_{n-1} are, therefore, $n-1$ and 0. The rank of J_{n-1} is 1. Hence the eigenvalues of $I-J_{n-1}/n$ are $1/n$ and 1. The determinant of each block is therefore $1/n$ and since there are $n+1$ blocks the result follows. □

A simple consequence of the lemma is that a set of MUBs are independent. Let us now turn to the general case. The diagonal blocks are the same and $(\gamma^{kk})^{-1}=(I-J_{n-1}/n)^{-1}=I+J_{n-1}$. Hence in block form we have

$$\det\Gamma(\mathcal{C})=\det\Gamma_0\cdot\det\begin{pmatrix} I & (I+J_{n-1})\gamma^{11} & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \cdots & I\end{pmatrix}. \tag{13}$$

That is, the off-diagonal blocks are multiplied by the matrix $I+J_{n-1}$. Consider γ^{kl}. Recall that $\gamma^{kl}(ij)=s^{kl}_{ij}-1/n,\ i,j\leqslant n-1$, where s^{kl}_{ij} are transition probabilities. An easy calculation shows that $(I+J_{n-1})\gamma^{kl}(ij)=s^{kl}_{ij}-s^{kl}_{in}$. The term s^{kl}_{in} appears because we omitted the nth basis vector from each measurement basis in the Hilbert space $\mathcal{H}$. If we had chosen another vector, say, the first then s^{kl}_{i1} would have been subtracted. The point is the information content depends on the *differences* of probabilities. Only in the case of MUBs are these differences all zero. Next we give an estimate in the general case.

Theorem 2. *Let* $\left|s^{kl}_{ij}-s^{kl}_{ir}\right|<\varepsilon$ *for some* $\varepsilon>0$. *Let* $\Gamma'=\Gamma_0^{-1}\Gamma(\mathcal{C})-I_{n^2-1}$ *and let* λ_m *be the minimum eigenvalue of* Γ'. *Suppose* $\lambda_m>-1$. *Then*

$$\mathrm{e}^{\frac{(n^2-n)^2(n^2-1)\varepsilon^2}{1+\lambda_m}}\frac{\det\Gamma(\mathcal{C})}{\det\Gamma_0}\geqslant 1. \tag{14}$$

Proof. The theorem is a consequence of an estimate given in [10]. From its definition the diagonal blocks $\Gamma(\mathcal{C})$ are positive semidefinite because it is a real matrix of the form $\langle \mathtt{b}_i,\mathtt{b}_j\rangle$ for vectors $\mathtt{b}_i$ in an appropriate dimension and its diagonal blocks are positive definite. Hence the estimate in [10] is applicable. The upper bound is just the Hadamard–Fischer inequality. The lower estimate in [10] is $\mathrm{e}^{\frac{-(n^2-1)\rho^2}{1+\lambda_m}}$, where $\rho = \max\{|\lambda_i| : \lambda_i \text{ an eigenvalue}\}$ is the spectral radius of Γ'. The fact that $\rho \leqslant \max\{|R_i|\}$, where $|R_i|$ is the sum of absolute values of the entries in the ith row of Γ' is a fact from linear algebra [11]. We get $n^2 - n$ because the diagonal blocks in Γ' are zero. □

As an illustration let $\varepsilon \leqslant 1/n^4$. Then a simple calculation yields $\det(\Gamma(\mathcal{C}))/\det(\Gamma_0) \geqslant \mathrm{e}^{-1/n^2}$. The corresponding loss in information is $O(1/n^2)$. In cases where MUBs are known to exist, that is, when n is a prime power, it is natural to expect that in some actual designing for testing in these bases there would be errors. If we can bound the errors by some ε then the information loss can be estimated. Even in cases where MUBs are not known to exist approximate MUBs may be constructed [12]. However, a direct application of the above estimates to their constructions does not yield very good lower bounds. If $\varepsilon \leqslant 1/n^3$, as in some cases of [12], then the information loss can be estimated to be less than $a = O(1)$. We now consider some generalizations of MUBs and give exact calculations for the corresponding measurement schemes.

5. Generalization of MUBs and calculation of determinants

We now consider a generalization of MUBs. Then we investigate the spectrum and the eigenvectors of the corresponding Γ. Our method provides an algorithm for solving equation (8) for the parameters characterizing the state. In these cases the method gives solutions even when the bases are complete but not necessarily independent. We apply the results and techniques developed in the first section to some bases constructed in [12]. We also give examples of the generalized MUBs.

5.1. A generalization of MUBs

In this section we consider the following generalization of mutually unbiased basis. Suppose that there exist bases $\mathcal{B}^i \equiv \left\{P^i_j = \left|\alpha^i_j\right\rangle\!\left\langle\alpha^i_j\right| : j = 1, \ldots, n\right\}$ for $i = 1, \ldots, r$ satisfying the following equations:

$$\mathtt{Tr}\left(P^i_j P^k_l\right) = s_{jl} \qquad \text{for all} \quad i, j, k, l. \tag{15}$$

We write S' for the $(n-1) \times (n-1)$ matrix with entries s_{jl}. We note that these equations express that for some predetermined orderings of the bases the matrices of transition probabilities $\mathcal{M}(\mathcal{B}^i, \mathcal{B}^k) = \left(\mathtt{Tr}\left(P^i_j P^k_l\right)\right)$ are all equal to S'. We observe that these matrices are doubly stochastic [14]. In the case of MUBs $S' = J_n/n$. Let S be the $(n-1) \times (n-1)$ matrix obtained by deleting the last row and column of S' and let $A = S - J_{n-1}/n$. The entries of A are scalar product (the trace product) of the operators $T^i_j = P^i_j - I_{n-1}/n$ and $T^k_l = P^k_l - I_{n-1}/n$, $1 \leqslant j, k \leqslant n-1$. We have noted earlier that the unknown state is completely determined if the probability distributions in the bases $\{\mathcal{B}^i\}$ are known provided the T^i_j span $l(\mathcal{H})$. The matrix connecting the probabilities and the state is the $(n-1)r \times (n-1)r$ matrix Γ of transition probabilities

$$\Gamma = \Gamma(\mathcal{C}) = \begin{pmatrix} \sigma & A & A & \cdots & A \\ A & \sigma & A & \cdots & A \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A & \cdots & \cdots & A & \sigma \end{pmatrix}.$$

We have to calculate the determinant of Γ. We do this by finding the eigenvalues of Γ. We will also indicate how to find the eigenvectors. First note that S must be a real symmetric matrix. If x is an eigenvalue corresponding to an eigenvector $\mathbf{X}$ of Γ then write the eigenvalue equation as

$$\Gamma\mathbf{X} \equiv \Gamma \begin{pmatrix} \mathtt{x}_1 \\ \mathtt{x}_2 \\ \vdots \\ \mathtt{x}_r \end{pmatrix} = x\mathbf{X} \tag{16}$$

where $\mathtt{x}_i$ are $(n-1)$-dimensional vectors. Then we have the equations

$$\sigma\mathtt{x}_i + A\sum_{j\neq i}\mathtt{x}_j = x\mathtt{x}_i \quad \Rightarrow \quad (xI_{n-1} - \sigma + A)\mathtt{x}_i = A\mathtt{y} \quad \text{with} \quad \mathtt{y} = \sum_{j=1}^{r}\mathtt{x}_j. \tag{17}$$

The last equation implies that the vectors $\mathtt{x}_i, i = 2, \ldots, r$ must be of the form $\mathtt{x}_i = \mathtt{x}_1 + \zeta_i$ where the vectors ζ_i are in the kernel of the operator $K(x) \equiv (xI_{n-1} - \sigma + A)$. It is easy to see that $K(x)\mathtt{u} = 0$ iff $S\mathtt{u}_i = (1-x)\mathtt{u}_i$. Therefore, the eigenvalues of Γ are of the form $1-d$, where d is an eigenvalue of S unless all the vectors $\mathtt{x}_i$ are equal. Let us first suppose this. That is, $\mathbf{X} = (\mathtt{z}, \mathtt{z}, \ldots, \mathtt{z})^T$. Then we must have

$$\left((r-1)S - \frac{r}{n}J_{n-1}\right)\mathtt{z} = G\mathtt{z} = (x-1)\mathtt{z}. \tag{18}$$

Since G is real symmetric we have $n-1$ eigenvectors with eigenvalues $1+d'$, where d' is an eigenvalue of G. We summarize the above discussion in the following.

Proposition 2. *Let Γ' be the matrix of transition probabilities among r bases written in a block form such that the (i,j)th block C^{ij} is the transition probabilities of the first $n-1$ vectors of the ith and jth bases, respectively. Suppose $C^{ij} = S$ if $i \neq j$ and $C^{ii} = I_{n-1}$. Let $\Gamma = \Gamma' - \frac{J_{(n-1)r}}{n}$. Then the eigenvalues of Γ are either $1+d'$ or $1-d$ where d' and d are the eigenvalues of $G = \left((r-1)S - \frac{r}{n}J_{n-1}\right)$ and S, respectively. The eigenvectors of Γ are of the form*

$$\mathtt{Z} = (\mathtt{z}, \mathtt{z}+\zeta_1, \mathtt{z}+\zeta_3, \ldots, \mathtt{z}+\zeta_{r-1})^T \tag{19}$$

where the vectors ζ_i are all zero (corresponding to eigenvalue $1+d'$) or are eigenvectors of S with the same eigenvalue.

The proposition gives the eigenvalues and the form of eigenvectors. To obtain multiplicities and to get the eigenvectors we have to know more about the structure of S. Let us call the eigenvalues and eigenvectors corresponding to d' and d type 1 and type 2, respectively. In general, suppose that none of the transition probabilities s_{ij} is zero. We call this an irreducible measurement scheme following a similar definition in the case of Markov chains [13]. For an irreducible measurement scheme only type 1 eigenvalues can be zero. For a type 2 eigenvalue can be zero if and only if the doubly substochastic matrix S has 1 as an eigenvalue. The above condition implies that

$$|S|_\infty \equiv \max_i \sum_j |s_{ij}| < 1.$$

That is, the matrix norm defined by taking the maximum of the sum of absolute values of the rows (see [15] chapter IV) is less than 1. If we had $\mathtt{x} = S\mathtt{x}$ then $\|\mathtt{x}\| \leqslant |S|_\infty\|\mathtt{x}\| < \|\mathtt{x}\|$ which is impossible unless $\mathtt{x} = 0$. Thus, Γ is singular if and only if the matrix G has an eigenvalue -1. Now let us turn our attention to G. First, recall that S is the submatrix of the $n \times n$ matrix S'

obtained by deleting the last row and column. We define $G' = (r-1)S' - \frac{r}{n}J_n$. Note that, like S', the matrix $S_D \equiv -G'$ has row and column sums equal to 1. However, the entries of S_D may be negative unless all the transition probabilities satisfy $s_{ij} \leqslant (n+1)/n^2$. Then, if all the entries of S_D are positive we conclude that the bases $\mathcal{B}^i$ are independent.

Now let us focus on type 2 eigenvectors. Thus, suppose $\mathtt{Z}$ is a type 2 eigenvector of Γ with eigenvalue x. It is given by equation (19) such that all the ζ_i are eigenvectors of S with eigenvalue $1-x$ and at least one of them is nonzero. Let $\zeta = \sum_i \zeta_i$. Then we have the condition

$$G\mathtt{z} = (x-1)(\mathtt{z}+\zeta) + \frac{\mathtt{Tr}(\zeta)}{n}\mathtt{e}. \tag{20}$$

Here the trace $\mathtt{Tr}$ of a vector is the sum of its entries and $\mathtt{e} = (1,\ldots,1)^t$. If we choose $\zeta = 0$ then the above equation reduces to $G\mathtt{z} = (x-1)\mathtt{z}$. This is possible only if $(x-1)$ is an eigenvalue of G. That is, S and G have eigenvalues $d = 1-x$ and $-d$, respectively. Suppose, there are m eigenvalues, counted with the multiplicities, for which this holds. Let $\{\zeta_1,\ldots,\zeta_m\}$ be the corresponding eigenvectors of S. Let $\mathcal{K}_i = \{(\mathtt{z}+\zeta_i, \mathtt{z}-\zeta_i, \mathtt{z},\ldots,\mathtt{z}), (\mathtt{z}+\zeta_i, \mathtt{z}, \mathtt{z}-\zeta_i, \mathtt{z},\ldots,\mathtt{z}), \ldots, (\mathtt{z}+\zeta_i, \mathtt{z},\ldots,\mathtt{z}, \mathtt{z}-\zeta_i)\}$. The set of vectors $\cup_i\mathcal{K}_i$ can be chosen to be linearly independent and in fact orthogonal. We have to be careful about the special case where S and G have common eigenvectors. It is not difficult to show that there are $(r-1)m$ such vectors. Next, if $(x-1)$ is *not* an eigenvalue of G then $G-(x-1)I_{n-1}$ is invertible and equation (20) has a unique solution for $\mathtt{z}$. Then we find $(r-1)(n-1-m)$ independent eigenvectors of the form $(\mathtt{z},\mathtt{z},\ldots,\mathtt{z}+\zeta,\mathtt{z},\ldots,\mathtt{z})$ for each eigenvector ζ of S belonging to the eigenvalue $1-x$. Add the $n-1$ type 1 eigenvectors. We have a complete set of $r(n-1)$ eigenvectors of Γ. We summarize the preceding analysis in the following theorem.

Theorem 3. *Let Γ' be the matrix of transition probabilities among r bases written in a block form such that the (i,j)th block C^{ij} is the transition probabilities of the first $n-1$ vectors of the ith and jth bases, respectively. Suppose $S^{ij} = S$ if $i \neq j$ and $S^{ii} = I_{n-1}$. Let $\Gamma = \Gamma' - \frac{J_{r(n-1)}}{n}$. Then the eigenvalues of Γ are either $1+d'$ (type 1) or $1-d$ (type 2) where d' and d are the eigenvalues of $G = \left((r-1)S - \frac{r}{n}J_{n-1}\right)$ and S, respectively. There are $n-1$ type 1 eigenvectors of Γ of the form $(\mathtt{z},\ldots,\mathtt{z})$ where $\mathtt{z}$ is an eigenvector of G. The type 2 eigenvectors are further divided into two classes.*

(1) The first class corresponds to those eigenvalues d for which $-d$ is an eigenvalue of G. If there are m such eigenvalues, counted with multiplicity, then we can construct $(r-1)m$ independent eigenvectors of Γ which are of the form $(\mathtt{z}+\zeta_i, \mathtt{z}-\zeta_i, \mathtt{z},\ldots,\mathtt{z})$ where $\zeta_i, i = 1,\ldots,r$ are eigenvectors of S.

(2) The second class of eigenvectors correspond to those eigenvalues d of S for which $-d$ not an eigenvalue of G. Then, the eigenvectors of Γ are of the form $(\mathtt{z}',\mathtt{z}',\ldots,\mathtt{z}'+\zeta,\mathtt{z}',\ldots,\mathtt{z}')^T$ where $\mathtt{z}' = (G+d)^{-1}(-d\zeta + (\mathtt{Tr}(\zeta)/n)\mathtt{e})$ and ζ an eigenvector of S. There are $(r-1)(n-1-m)$ independent vectors in this class.

The theorem gives a concrete way of constructing a complete set of eigenvectors of Γ. This set of eigenvectors need not be orthogonal. But we can construct a complete orthonormal set by the Gramm–Schmidt procedure. Let U be the matrix of an *orthonormal* set of eigenvectors. Since $U^\dagger\Gamma U = D$ is the diagonal matrix of eigenvalues of Γ we get $\Gamma^{-1} = UD^{-1}U^\dagger$. We take the generalized inverse of Γ when the bases are complete but not independent. We have 'solved' the state reconstruction problem. In reality, we do not get the probabilities but only the *frequencies* and we have to use some statistical estimation method for the probabilities. As a consequence of the theorem we get the following corollary.

Corollary 1

$$\det(\Gamma) = (\det(I - S))^{r-1}(\det(I + G)).$$

All the determinant calculations in the preceding sections can be derived from the above corollary. An immediate consequence of the above theorem is the following corollary which deals with a somewhat more general case. We use the notation of the theorem.

Corollary 2. *Suppose we have r bases $\{\mathcal{B}_1, \ldots, \mathcal{B}_r\}$ such that the each of the first k bases is mutually unbiased with respect to the rest. Suppose further that the remaining $r - k$ bases have a constant transition probability matrix S.*

Then there are three sets of eigenvalues for Γ. The first, called type 0, contains only 1 and $1/n$ with multiplicity $(n - 2)k$ and k, respectively. The second and third types are the type 1 and type 3 eigenvalues of the theorem. Their multiplicities are obtained by replacing r with $r - k$ in the theorem. The eigenvectors for type 0 are $(1, \ldots, 1, \mathbf{0})^T$ (eigenvalue 1) and $(1, -1, 1, \ldots, 1, \mathbf{0})^T, \ldots, (1, 1, \ldots, 1, -1, \mathbf{0})^T$, where $\mathbf{0}$ is the zero vector in dimension $(r - k)(n - 1)$. The type 1 and type 2 eigenvectors are of the form $(\mathbf{0}, \mathbf{X})$ and $(\mathbf{0}, \mathbf{Z})$, where $\mathbf{0}$ is now the zero vector in dimension $k(n - 1)$ and $\mathbf{X}$ and $\mathbf{Z}$ are the type 1 and type 2 eigenvectors given in the theorem with r replaced by $r - k$. Finally, $\det(\Gamma) = (\det(I - S))^{r-k-1} \det(I + G)/n^k$.

The corollary is immediate because of the structure of Γ. In the first k rows (and columns) the off-diagonal blocks are zero and in the rest the off-diagonals blocks are all equal to $S - J_{n-1}/n$.

We conclude this section with several examples.

5.2. Example 1

In [12] the authors use some number theoretic results on character sums to define sets of bases which approximate MUBs. We only consider the case where the dimension $n = p - 1$, p a prime. Let F_p be the finite field of integers modulo p and $F_p^{\times}$ be the multiplicative group of nonzero elements. It is well known that $F_p^{\times}$ is cyclic, that is, there is an element $u \in F_p^{\times}$ such that $F_p^{\times} = \{u^i, i = 1, \ldots, n\}$. Let $\chi : F_p^{\times} \to \mathbb{C}$ be the function $\chi(u) = \mathrm{e}^{2\pi\, \mathrm{i}u/n}$ and $\chi(u^j) = \chi^j(u)$. The function χ is a *character* on the group $F_p^{\times}$. Let

$$\left(\left|\alpha_k^j\right\rangle\right)_l = \frac{1}{\sqrt{n}} \mathrm{e}^{2\pi i j l/p} \chi^k(l), \qquad 0 \leqslant j \leqslant n-1, \quad j, k = 1, \ldots, n \quad \text{and} \quad \mathcal{B}_j = \left\{\left|\alpha_k^j\right\rangle\right\}.$$

Here $\left(\left|\alpha_k^j\right\rangle\right)_l$ is the lth element of the complex column vector $\left|\alpha_k^j\right\rangle$. The bases are $\{\mathcal{B}_0, \ldots, \mathcal{B}_n\}$ with $\mathcal{B}_0 = \left\{\left|\alpha_k^0\right\rangle\right\}$ the standard basis. In [12] an estimate of the product $\left|\left\langle\alpha_i^a\middle|\alpha_j^b\right\rangle\right|$ is given. However, an exact calculation is possible[3] in this case, yielding for a or $b \neq 0$,

$$\begin{aligned} \left|\left\langle\alpha_i^a\middle|\alpha_j^b\right\rangle\right|^2 = s_{ij}^{ab} &= \delta_{ij}, && a = b \\ &= \frac{n+1}{n^2}, && a \neq b, i \neq j \\ &= \frac{1}{n^2}, && a \neq b, i = j. \end{aligned} \tag{21}$$

[3] This fact was pointed out to me by Professor Igor Shparlinski. I am grateful to him for several illuminating discussions.

Clearly for a or $b = 0$ (but not both), $s_{ij}^{ab} = 1/n$. The basis $\mathcal{B}_0$ is unbiased with respect to the rest. The matrices S and G are given by

$$S = -\frac{1}{n} I_{n-1} + \frac{n+1}{n^2} J_{n-1} \qquad \text{and} \qquad G = -\frac{n-1}{n} I_{n-1} - \frac{1}{n^2} J_{n-1}.$$

The eigenvalues of S are $-1/n$ and $1-(n+1)/n^2$ with respective multiplicities $n-2$ and 1 and eigenvalues of G are $1/n-1$ and $1/n^2-1$ with the same multiplicities. Using corollary 2 the eigenvalues of Γ with multiplicities are as follows: $(1, n-2)$, $(1/n, 1)$ (type 0), $(1/n, n-2)$, $(1/n^2, 1)$ (type 1), $(1+1/n, (n-1)(n-2))$, $((n+1)/n^2, (n-1))$ (type 2). Hence the determinant is

$$\det\Gamma = \left(\frac{(n+1)^{n-1}}{n^n}\right)^{n-1} \frac{1}{n^{n+1}}.$$

Using lemma 1 we see that the information loss in measuring in the KRSW with respect to a corresponding MUBs is proportional to $\Delta\mathcal{I} = \log(\det\Gamma)_0 - \log(\det\Gamma) = (n-1)(n\log n - (n-1)\log(n+1))$. Hence $(n-1)(\log(n+1) - \log e) < \Delta\mathcal{I} < (n-1)(\log(n+1) - \log 2)$. The base of the logarithm is arbitrary here although it is usually taken to be 2 in information theory. We note that information loss is of the same order as the information gain of MUBs.

Using corollary 2 we can also find out the eigenvectors of Γ. The calculations are straightforward and we only sketch the construction for some of the type 2 eigenvectors corresponding to the eigenvalue $d = 1-(n+1)/n^2$. Since $-d$ is not an eigenvalue of G the eigenvectors are given by the second form of type 2 vectors in the corollary (and the preceding theorem). The corresponding eigenvector ζ of S may be taken to be $\mathtt{e}$. We have $(G+d)^{-1}(-d\zeta + \mathtt{Tr}\,\zeta/n\mathtt{e} = -\mathtt{e}/n$. Hence, the eigenvectors of Γ for eigenvalue $1-d$ are $(\mathtt{e}/n, -(1-1/n)\mathtt{e}, \mathtt{e}, \ldots, \mathtt{e})$, $(\mathtt{e}/n, \mathtt{e}/n, -(1-1/n)\mathtt{e}, \ldots, \mathtt{e}/n)$, $(\mathtt{e}/n, \mathtt{e}/n, \ldots, -(1-1/n)\mathtt{e})$. We can similarly construct eigenvectors. In fact, we can get an orthonormal set of eigenvectors yielding a solution to the state determination problem. Finally we observe that if we had a basis $\mathcal{B}_0'$ (instead of $\mathcal{B}_0$) such that *all* the transition probability matrices were equal, then the determinant of the corresponding Γ is zero and the bases are dependent.

5.3. *Example 2 (dimension 2)*

$$\mathcal{B}^1 = \left\{\alpha_0^1 = \begin{pmatrix}1\\0\end{pmatrix}, \alpha_0^1 = \begin{pmatrix}0\\1\end{pmatrix}\right\}, \qquad \mathcal{B}^2 = \left\{\alpha_0^2 = \begin{pmatrix}a\\b\end{pmatrix} \alpha_1^2 = \begin{pmatrix}b\\-a\end{pmatrix}\right\} \qquad \text{and}$$

$$\mathcal{B}^3 = \left\{\alpha_0^3 = \begin{pmatrix}a\\be^{it}\end{pmatrix} \alpha_1^3 = \begin{pmatrix}b\\-a\,\mathrm{e}^{\mathrm{i}t}\end{pmatrix}\right\},$$

$$\cos t = 1 - \frac{1}{2a^2} \qquad \text{and} \qquad a = \sqrt{1-b^2} \geqslant 1/2.$$

The transition probability matrix for any pair of bases is

$$\begin{pmatrix} a^2 & b^2 \\ b^2 & a^2 \end{pmatrix}.$$

Calling this collection of bases $\mathcal{C}$ we obtain the corresponding

$$\Gamma(\mathcal{C}) = \begin{pmatrix} \frac{1}{2} & a^2 - \frac{1}{2} & a^2 - \frac{1}{2} \\ a^2 - \frac{1}{2} & \frac{1}{2} & a^2 - \frac{1}{2} \\ a^2 - \frac{1}{2} & a^2 - \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

In this simple case the blocks of Γ are simply numbers (one-dimensional). The eigenvalue of type 1 is $2a^2 - 1/2$ with one eigenvector $(1, 1, 1)^T$ and the type 2 eigenvalue $1 - a^2$ has two eigenvectors $(1, -2, 1)^T$ and $(1, 1, -2)^T$, and the determinant is $(2a^2 - 1/2)(1 - a^2)^2$.

5.4. Example 3 (dimension 3)

In the previous example in two dimensions the transition probability matrix was the same for any two bases. This is the situation that theorem 3 deals with. However, the first example draws on corollary 2. In three dimensions there is no 'easy' example corresponding to the theorem although some rather messy calculations suggest that such bases should exist. But it is relatively easier to find examples corresponding to the corollary. Thus let ω be a cube root of 1 and define

$$U_1 = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & 1\\ 1 & \omega^2 & \omega\\ 1 & \omega & \omega^2\end{pmatrix} \qquad U_2 = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & 1\\ \mathrm{e}^{\mathrm{i}\alpha} & \mathrm{e}^{\mathrm{i}\alpha}\omega^2 & \mathrm{e}^{\mathrm{i}\alpha}\omega\\ 1 & \omega & \omega^2\end{pmatrix}$$
$$U_3 = \frac{1}{\sqrt{3}}\begin{pmatrix}1 & 1 & 1\\ 1 & \omega^2 & \omega\\ \mathrm{e}^{-\mathrm{i}\alpha} & \mathrm{e}^{-\mathrm{i}\alpha}\omega & \mathrm{e}^{-\mathrm{i}\alpha}\omega^2\end{pmatrix}.$$

We also let U_0 be the identity matrix. Our bases consist of the column vectors of these matrices. Let K_{ij} denote the transition probability matrix between bases i and j, respectively. Then

$$P_{0i} = \frac{1}{3}\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix} \qquad P_{ij} = S' = \frac{1}{9}\begin{pmatrix}5+4\cos\alpha & 2(1-\cos\alpha) & 2(1-\cos\alpha)\\ 2(1-\cos\alpha) & 5+4\cos\alpha & 2(1-\cos\alpha)\\ 2(1-\cos\alpha) & 2(1-\cos\alpha) & 5+4\cos\alpha\end{pmatrix}.$$

We write the matrix S and G in a convenient form

$$S = \frac{1}{3}\begin{pmatrix}a & b\\ b & a\end{pmatrix} \qquad G = \begin{pmatrix}\frac{2a}{3}-1 & \frac{2b}{3}-1\\ \frac{2b}{3}-1 & \frac{2a}{3}-1\end{pmatrix} \qquad a = 3-2b = \frac{5+4\cos\alpha}{3}$$

$$\text{and} \qquad \Gamma = \begin{pmatrix}I_2 - J_2/3 & 0 & 0 & 0\\ 0 & I_2 - J_2/3 & S - J_2/3 & S - J_2/3\\ 0 & S - J_2/3 & I_2 - J_2/3 & S - J_2/3\\ 0 & S - J_2/3 & S - J_2/3 & I_2 - J_2/3\end{pmatrix}.$$

The eigenvalues of S and G are $(a \pm b)/3$ and $\{2(a+b)/3-2, 2(a-b)/3\}$, respectively. Hence the eigenvalues of Γ are $\{1, 1/3\}$ (type 0), $\{2(a+b)/3-1, 1+2(a-b)/3\}$ (type 1), $\{1-(a+b)/3, 1-(a-b)/3\}$ (type 2). Next we list the eigenvectors. First we observe that there is no eigenvalue d of S such that $-d$ is eigenvalue of G unless $b = 4/3$, that is, $\cos\alpha = -1$. We omit this case as it is easy to handle. Hence, only the second construction of type 2 eigenvectors applies. The eigenvectors of S (and G) are $\beta_1 = (1, 1)^T$ and $\beta_2 = (1, -1)^T$. We can now list all eight eigenvectors of Γ:

`type 0:`$(\beta_1, 0, 0, 0)^T, (\beta_2, 0, 0, 0)^T$ `type 1:`$(0, \beta_1, \beta_1, \beta_1)^T, (0, \beta_2, \beta_2, \beta_2)^T$
`type 2:`$(0, \beta_1, -2\beta_1, \beta_1)^T, (0, \beta_1, \beta_1, -2\beta_1)^T, (0, \beta_2, -2\beta_2, \beta_2)^T$, and
$(0, \beta_2, \beta_2, -2\beta_2)^T$.

The determinant calculation is easy. It is also easy to find an orthogonal set of eigenvectors from the above and we can compute the inverse of Γ at little cost. We have the formal solution to the state determination problem. It should be noted that in this example $a = b = 1$ corresponds to MUBs.

6. Discussion

We analyzed the information content of a quantum state determination experiment using projective measurements in different bases. The information gain is a functional of the prior distribution. We get the average information gain by averaging over the initial distribution. When the number of measurements is large and the prior distribution is uniform, the information gain is proportional to the log of volume of the parallelepiped spanned by the basis vectors after an affine translation. The calculation of the determinant in the case of MUBs and some generalization yields rich dividends. In all these cases we obtain formal solutions to the problem of state determination when the measurement scheme is complete. The methods for calculating the eigenvalues could be extended to more general bases. We also give an estimate of the information gain in the general case. Characterization of optimal measurements in the general case is a difficult problem. Three other difficult problems are (1) estimating information gain in measurements on infinite dimensional systems, (2) incomplete measurements and (3) information gain for priors other than the uniform distribution. Yet another important issue is the *existence* of bases of the preceding sections. The KRSW construction provides some examples. More such constructions may be carried out using *character sums*—a very rich and fertile area in number theory [18]. I aim to address these problems in the future.

Acknowledgment

I wish to thank Professor William Wootters for kindly sending me his thesis.

References

[1] Rehácek J, Englert B-G and Kaszlikowski D 2004 *Phys. Rev.* A **70** 052321
[2] Wootters W K and Fields B D 1989 *Ann. Phys., NY* **191** 363
[3] Helstrom C W *Quantum Estimation and Detection Theory* (New York: Academic)
[4] Hayashi M 1998 *J. Phys. A: Math. Gen.* **31** 4633
[5] Ivanovic I D 1981 *J. Phys. A: Math. Gen.* **14** 3241
[6] Lindley D V 1956 *Ann. Math. Stat.* **27** 986
[7] Khinchin A I 1957 *Mathematical Foundations of Information Theory* (New York: Dover)
[8] Gnedenko B V 1967 *The Theory of Probability* (New York: Chelsea) pp 95–117
[9] Engel G and Schneider H 1976 *Linear Multilinear Algebra* **4** 155
[10] Ipsen I C F and Lee D J 2003 Determinant approximation (http://www.ncsu.edu/ipsen/)
[11] Marcus M and Minc H 1992 *A Survey of Matrix Theory and Matrix Inequalities* (New York: Dover)
[12] Klappenecker A, Rotteler M, Shparlinski I E and Winterhof A 2005 *J. Math. Phys.* **46** 082104
[13] Feller W 1968 *An Introduction to Probability Theory and Its Applications* (New York: Wiley) chapter XV
[14] Ando T 1989 *Linear Algebra Appl.* **118** 163–88
[15] Bhatia R 1997 *Matrix Analysis* (Berlin: Springer)
[16] Penrose R 1955 *Proc. Camb. Phil. Soc.* **51** 406–13
[17] Horn R A and Johnson C R 1999 *Matrix Analysis* (Cambridge: Cambridge University Press)
[18] Shparlinski I *Lecture Notes* (http://www.ma.utexas.edu/users/voloch/expsums.html)
[19] Hall M J W 1998 *Phys. Lett.* A **242** 123–9
[20] Slater P B 1998 *Phys. Lett.* A **247** 1–8
[21] Srednicki M 1998 *Phys. Rev.* A **71** 052107
[22] Balasubramanian V 1998 *Neural Comput.* **9** 349