Quantum parameter estimation of a generalized Pauli channel

2003 J. Phys. A: Math. Gen. 36 8093

(http://iopscience.iop.org/0305-4470/36/29/314)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **36** (2003) 8093–8103 PII: S0305-4470(03)63232-4

Quantum parameter estimation of a generalized Pauli channel

Akio Fujiwara and Hiroshi Imai

Department of Mathematics, Osaka University, Toyonaka, Osaka 560-0043, Japan

E-mail: fujiwara@math.wani.osaka-u.ac.jp

Received 9 May 2003, in final form 6 June 2003
Published 8 July 2003
Online at stacks.iop.org/JPhysA/36/8093

Abstract

We present a quantum parameter estimation theory for a generalized Pauli channel $\Gamma_\theta : \mathcal{S}(\mathbb{C}^d) \to \mathcal{S}(\mathbb{C}^d)$, where the parameter θ is regarded as a coordinate system of the probability simplex $\mathcal{P}^{d^2-1}$. We show that for each degree n of extension $(\mathrm{id} \otimes \Gamma_\theta)^{\otimes n} : \mathcal{S}((\mathbb{C}^d \otimes \mathbb{C}^d)^{\otimes n}) \to \mathcal{S}((\mathbb{C}^d \otimes \mathbb{C}^d)^{\otimes n})$, the SLD Fisher information matrix for the output states takes the maximum when the input state is an n-tensor product of a maximally entangled state $\tau^{ME} \in \mathcal{S}(\mathbb{C}^d \otimes \mathbb{C}^d)$. We further prove that for the corresponding quantum Cramér–Rao inequality, there is an efficient estimator if and only if the parameter θ is ∇^m-affine in $\mathcal{P}^{d^2-1}$. These results rely on the fact that the family $\{\mathrm{id} \otimes \Gamma_\theta(\tau^{ME})\}_\theta$ of output states can be identified with $\mathcal{P}^{d^2-1}$ in the sense of quantum information geometry. This fact further allows us to investigate submodels of generalized Pauli channels in a unified manner.

PACS numbers: 03.67.−a, 03.65.Ud, 89.70.+c

1. Introduction

Since almost every quantum protocol assumes *a priori* knowledge of the behaviour of the quantum channel under consideration, there is no doubt that identifying the channel is of fundamental importance in quantum information theory. It is, however, not very long since the quantum channel identification problem was directed proper attention, and the theory of finding an optimal estimation scheme has not been investigated so far, with only a few exceptions [1–4]. This paper addresses the problem of finding an optimal estimation scheme for a generalized Pauli channel, based on noncommutative parameter estimation theory [5, 6] and information geometry [7]. In view of applications, generalized Pauli channels form a reasonably large class of quantum channels, including many important submodels such as the bit flip channel, the phase damping channel and the depolarizing channel [8].

Let $\mathcal{S}(\mathcal{H})$ be the set of density operators on a Hilbert space $\mathcal{H}$. A Pauli channel $\Gamma : \mathcal{S}(\mathbb{C}^2) \to \mathcal{S}(\mathbb{C}^2)$ acting on a two-dimensional quantum system is defined by

$$\Gamma(\tau) = \sum_{i=0}^{3} p_i(\sigma_i \tau \sigma_i^*)$$

where $p = (p_0, p_1, p_2, p_3)$ is a probability vector, $\sigma_0 = I$, and $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices. For mathematical simplicity, we assume that the probability vector p is strictly positive ($p > 0$), i.e., $p_i > 0$ for all i. Obviously, there is a one-to-one correspondence between the set of Pauli channels and the three-dimensional open probability simplex:

$$\mathcal{P}^3 = \left\{ p = (p_0, p_1, p_2, p_3); \forall p_i > 0, \sum_{i=0}^{3} p_i = 1 \right\}.$$

A generalized Pauli channel $\Gamma : \mathcal{S}(\mathbb{C}^d) \to \mathcal{S}(\mathbb{C}^d)$ acting on a d-dimensional quantum system is defined in a similar way [9, 10]:

$$\Gamma(\tau) = \sum_{i,j=0}^{d-1} p_{ij}(\sigma_{ij} \tau \sigma_{ij}^*)$$

where $p = (p_{ij})_{0 \leqslant i,j \leqslant d-1}$ is a probability vector that belongs to the $(d^2 - 1)$-dimensional open probability simplex $\mathcal{P}^{d^2-1}$, and σ_{ij} are the unitary operators that act on the standard basis $\{e_\ell\}_{1 \leqslant \ell \leqslant d}$ of $\mathbb{C}^d$ as

$$\sigma_{ij} e_\ell = \omega^{i\ell} e_{j+\ell}.$$

Here ω is the primitive dth root of unity, and the subscript of the basis is understood modulo d. Note that

$$\operatorname{Tr} \sigma_{ij}^* \sigma_{i'j'} = d\delta_{ii'}\delta_{jj'}.$$

Letting $\theta = (\theta^\lambda)_{1 \leqslant \lambda \leqslant d^2-1}$ be a global coordinate system of the probability simplex $\mathcal{P}^{d^2-1}$, one can specify a generalized Pauli channel Γ by the corresponding coordinate θ, whereby we can denote $\Gamma = \Gamma_\theta$. For example, the components of probability vector $p = (p_{ij})$, with p_{00} removed (since $p_{00} = 1 - \sum_{(i,j) \neq (0,0)} p_{ij}$), form a global ∇^m-affine coordinate system of $\mathcal{P}^{d^2-1}$. Any other ∇^m-affine coordinate system of $\mathcal{P}^{d^2-1}$ is obtained by a regular affine transformation of this coordinate system [7]. For notational simplicity, we identify the pair of indices (i, j) with the integer $k := di + j$, and use it as a new index:

$$\Gamma_\theta(\tau) = \sum_{k=0}^{d^2-1} p_k(\theta)(\sigma_k \tau \sigma_k^*). \tag{1}$$

We are interested in estimating the true value of the parameter θ, given an unknown (generalized) Pauli channel Γ_θ. The general scheme of estimating an unknown quantum channel is as follows [1, 3]: one inputs a well-prepared state $\tau^{(n)} \in \mathcal{S}((\mathbb{C}^d \otimes \mathbb{C}^d)^{\otimes n})$ to the extended channel $(\mathrm{id} \otimes \Gamma_\theta)^{\otimes n}$ and estimate the parameter θ by applying a certain measurement $M^{(n)}$ on the output state $(\mathrm{id} \otimes \Gamma_\theta)^{\otimes n}(\tau^{(n)})$. The problem thus amounts to finding an optimal pair of input $\tau^{(n)}$ and (unbiased) estimator $M^{(n)}$ for the parameter θ.

Let $J_\theta^{(n)}(\tau^{(n)})$ be the symmetric logarithmic derivative (SLD) Fisher information matrix of the output states $(\mathrm{id} \otimes \Gamma_\theta)^{\otimes n}(\tau^{(n)})$. For notational simplicity, the superscript (n) representing the degree of extension will be omitted when $n = 1$. The main result of this paper is the following.

Theorem 1. *For all* $\tau^{(n)} \in \mathcal{S}((\mathbb{C}^d \otimes \mathbb{C}^d)^{\otimes n})$*, it holds that*

$$J_\theta^{(n)}(\tau^{(n)}) \leqslant J_\theta^{(n)}((\tau^{ME})^{\otimes n}) = nJ_\theta(\tau^{ME})$$

where $\tau^{ME} \in \mathcal{S}(\mathbb{C}^d \otimes \mathbb{C}^d)$ *is an arbitrary maximally entangled state. As regards the corresponding SLD Cramér–Rao inequality*

$$nV_\theta[M^{(n)}] \geqslant J_\theta(\tau^{ME})^{-1}$$

there is an efficient estimator $M^{(n)}$ *for* θ *that uniformly achieves the lower bound if and only if the parameter* θ *is* ∇^m*-affine in* $\mathcal{P}^{d^2-1}$.

Theorem 1 implies that the optimal strategy for estimating a generalized Pauli channel is of iid type: one need only repeat the efficient parameter estimation scheme for $\mathrm{id} \otimes \Gamma_\theta(\tau^{ME})$, since this strategy achieves the most informative lower bound $J_\theta(\tau^{ME})^{-1}$ for all degrees n of extension. In other words, one cannot acquire more information than $J_\theta(\tau^{ME})$ per extension even if one invokes other entangled inputs and/or collective measurements that straddle (possibly) larger Hilbert spaces.

As will be clarified in section 2, theorem 1 essentially relies on the fact that the family $\{\mathrm{id} \otimes \Gamma_\theta(\tau^{ME})\}_\theta$ of output states can be identified with $\mathcal{P}^{d^2-1}$ in the sense of quantum information geometry. This observation further allows us to investigate submodels of generalized Pauli channels in a unified manner. For example, depolarizing channels form a ∇^e-geodesic in $\mathcal{P}^{d^2-1}$, and its ∇^m-affine parameter (such as the magnitude of depolarization) has an efficient estimator. This settles the open problems posed in [1] and [4].

The paper is organized as follows. Section 2 is devoted to the proof of the main theorem. Section 3 explores a systematic treatment of parameter estimation for submodels of generalized Pauli channels, based on information geometry. Further discussions and remarks are presented in section 4. In order for the exposition to be reasonably self-contained, we present a brief account of quantum information geometry in the appendix.

2. Proof of main theorem

Let $\mathcal{H}$ be a D-dimensional Hilbert space, and let $\Lambda_\theta : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})$ be a smooth parametric family of quantum channels defined by

$$\Lambda_\theta(\tau) = \sum_{k=1}^{K} p_k(\theta) U_k \tau U_k^* \tag{2}$$

where U_k are fixed unitary operators that satisfy $\mathrm{Tr}\, U_k^* U_\ell = 0$ for $k \neq \ell$, and $p(\theta) = (p_k(\theta))_{1 \leqslant k \leqslant K}$ is a probability measure parametrized by $\theta = (\theta^\lambda)$. We assume that $p_k(\theta) > 0$ for all k and θ. Let us explore the estimation theory for the parameter θ in which we make use of the extension $\mathrm{id} \otimes \Lambda_\theta : \mathcal{S}(\mathcal{H} \otimes \mathcal{H}) \to \mathcal{S}(\mathcal{H} \otimes \mathcal{H})$.

Given an arbitrary unit vector $\psi \in \mathcal{H} \otimes \mathcal{H}$, there are orthonormal bases $(e_i)_{1 \leqslant i \leqslant D}$ and $(f_j)_{1 \leqslant j \leqslant D}$ of $\mathcal{H}$ and a probability vector $\alpha = (\alpha_i)_{1 \leqslant i \leqslant D}$ such that

$$\psi = \sum_{i=1}^{D} \sqrt{\alpha_i}\, e_i \otimes f_i. \tag{3}$$

This is sometimes referred to as a Schmidt decomposition. When we are concerned only with the dependence of some property on the degree α of entanglement, we will not specify the Schmidt bases $(e_i)_i$ and $(f_i)_i$. For example, we denote the vector (3) by ψ^α and the corresponding pure state $|\psi^\alpha\rangle\langle\psi^\alpha|$ by τ^α. A pure state $\tau^u = |\psi^u\rangle\langle\psi^u|$ that corresponds to

the uniform distribution $u := (1/D, \ldots, 1/D)$ is called a maximally entangled state. Given an input pure state τ^α, let $\rho_\theta^\alpha := \mathrm{id} \otimes \Lambda_\theta(\tau^\alpha)$ for notational simplicity.

Lemma 2. *The family* $\left\{\rho_\theta^u\right\}_\theta$ *is commutative.*

Proof. The output state ρ_θ^u is explicitly written as

$$\rho_\theta^u = \frac{1}{D}\sum_{i,j=1}^{D} |e_i\rangle\langle e_j| \otimes \Lambda_\theta(|f_i\rangle\langle f_j|) = \sum_{k=1}^{K} p_k(\theta)|g_k\rangle\langle g_k| \tag{4}$$

where

$$g_k := \frac{1}{\sqrt{D}}\sum_{i=1}^{D} e_i \otimes (U_k f_i).$$

Since $\mathrm{Tr}\, U_k^* U_\ell = D\delta_{k\ell}$, the vectors $\{g_k\}_k$ are orthonormal. Thus (4) gives a simultaneous spectral decomposition of the family $\left\{\rho_\theta^u\right\}_\theta$. □

It follows from lemma 2 that the SLD and the right logarithmic derivative (RLD) of $\left\{\rho_\theta^u\right\}_\theta$ are identical.

Lemma 3. *For* $\alpha > 0$*, the RLD Fisher information matrix of the output family* $\left\{\rho_\theta^\alpha\right\}_\theta$ *is independent of* α*, and of the Schmidt bases* $(e_i)_i$ *and* $(f_i)_i$*.*

Proof. The output state ρ_θ^α is rewritten as

$$\rho_\theta^\alpha = A^\alpha \rho_\theta A^\alpha$$

where $\rho_\theta := \rho_\theta^u$ and

$$A^\alpha := \sqrt{D}\sum_{i=1}^{D} \sqrt{\alpha_i}|e_i\rangle\langle e_i| \otimes I.$$

The A^α is invertible for $\alpha > 0$, and the RLD with respect to the λth parameter θ^λ is given by

$$L_{\theta,\lambda}^\alpha := \left(\rho_\theta^\alpha\right)^{-1}\left(\partial_\lambda \rho_\theta^\alpha\right) = (A^\alpha)^{-1}(\rho_\theta)^{-1}(\partial_\lambda \rho_\theta)A^\alpha$$

where $\partial_\lambda := \partial/\partial\theta^\lambda$. The (λ, μ)th entry of the RLD Fisher information matrix $J_\theta^R(\tau^\alpha)$ for the input τ^α then becomes

$$\left(J_\theta^R(\tau^\alpha)\right)_{\lambda\mu} := \mathrm{Tr}\,\rho_\theta^\alpha L_{\theta,\mu}^\alpha \left(L_{\theta,\lambda}^\alpha\right)^* = \mathrm{Tr}(A^\alpha)^2(\partial_\lambda \rho_\theta)\rho_\theta^{-1}(\partial_\mu \rho_\theta).$$

Now by using the spectral decomposition (4), we have

$$(\partial_\lambda \rho_\theta)\rho_\theta^{-1}(\partial_\mu \rho_\theta) = \sum_k \frac{\partial_\lambda p_k(\theta)\partial_\mu p_k(\theta)}{p_k(\theta)}|g_k\rangle\langle g_k|.$$

Further it is easily verified that for each k

$$\mathrm{Tr}(A^\alpha)^2|g_k\rangle\langle g_k| = \sum_{i=1}^{D}\alpha_i = 1.$$

Thus we obtain

$$\left(J_\theta^R(\tau^\alpha)\right)_{\lambda\mu} = \sum_k \frac{\partial_\lambda p_k(\theta)\partial_\mu p_k(\theta)}{p_k(\theta)}.$$

This is nothing but the classical Fisher information of $p(\theta)$, which will be denoted by $\hat{J}_\theta$. In particular, it is independent of α and the Schmidt bases. □

Lemma 4. *For each θ, the SLD Fisher information matrix $J_\theta(\tau)$ of the output states* $\mathrm{id}\otimes\Lambda_\theta(\tau)$ *takes the maximum $\hat{J}_\theta$ when the input τ is a maximally entangled pure state τ^u.*

Proof. For a pure state input $\tau = \tau^\alpha$ with $\alpha > 0$, we have

$$J_\theta(\tau^\alpha) \leqslant J_\theta^R(\tau^\alpha) = J_\theta^R(\tau^u) = J_\theta(\tau^u).$$

Here the inequality follows from the well-known fact that the SLD Fisher metric is not greater than the real part of the RLD Fisher metric (see, for instance, [11]), the next equality from lemma 3, and the last equality from lemma 2. Further, by a continuity argument, the above established inequality

$$J_\theta(\tau^\alpha) \leqslant J_\theta(\tau^u) \quad (= \hat{J}_\theta) \tag{5}$$

holds for any probability vector α.

For a generic input $\tau \in \mathcal{S}(\mathcal{H}\otimes\mathcal{H})$, let

$$\tau = \sum_i q_i \tau_i^{\alpha_i} \qquad \left(\forall q_i > 0, \sum_i q_i = 1\right)$$

be a pure state decomposition. Then the convexity of the SLD Fisher metric [1] and the inequality (5) show that

$$J_\theta(\tau) \leqslant \sum_i q_i J_\theta\left(\tau_i^{\alpha_i}\right) \leqslant J_\theta(\tau^u)$$

as desired. □

Proof of theorem 1. Let Γ_θ be the generalized Pauli channel defined by (1). By a suitable rearrangement of the constituent Hilbert spaces $\mathbb{C}^d$, we identify $(\mathrm{id}\otimes\Gamma_\theta)^{\otimes n}$ with $(\mathrm{id})^{\otimes n}\otimes(\Gamma_\theta)^{\otimes n}$. The extended channel $(\Gamma_\theta)^{\otimes n}$ acts on $\mathcal{S}((\mathbb{C}^d)^{\otimes n})$ as

$$(\Gamma_\theta)^{\otimes n}(\tau) = \sum_{i_1,\ldots,i_n} p_{i_1,\ldots,i_n}(\theta)\sigma_{i_1,\ldots,i_n}\tau\sigma_{i_1,\ldots,i_n}^*$$

where $p_{i_1,\ldots,i_n}(\theta) := p_{i_1}(\theta)\cdots p_{i_n}(\theta)$ is the iid extension of the probability measure $p(\theta)$, and $\sigma_{i_1,\ldots,i_n} := \sigma_{i_1}\otimes\cdots\otimes\sigma_{i_n}$. Note that $\sigma_{i_1,\ldots,i_n}$ are unitary operators that satisfy

$$\mathrm{Tr}\,\sigma_{i_1,\ldots,i_n}^*\sigma_{j_1,\ldots,j_n} = d^n\delta_{i_1 j_1}\cdots\delta_{i_n j_n}.$$

Thus the channel $\Lambda_\theta := (\Gamma_\theta)^{\otimes n}$ is of the type (2) with $\mathcal{H} = (\mathbb{C}^d)^{\otimes n}$.

As we have rearranged the constituent Hilbert spaces, the iid extension $(\tau^{ME})^{\otimes n}$ of a maximally entangled state τ^{ME} on $\mathbb{C}^d\otimes\mathbb{C}^d$, on which $\mathrm{id}\otimes\Gamma_\theta$ acts, is also a maximally entangled state on $\mathcal{H}\otimes\mathcal{H}$ on which $(\mathrm{id})^{\otimes n}\otimes(\Gamma_\theta)^{\otimes n}$ acts, in that

$$\left(\frac{1}{d}\sum_i\sum_j |e_i\rangle\langle e_j|\otimes|f_i\rangle\langle f_j|\right)^{\otimes n} \sim \frac{1}{d^n}\sum_{i_1,\ldots,i_n}\sum_{j_1,\ldots,j_n}\left|e_{i_1}\otimes\cdots\otimes e_{i_n}\right\rangle\left\langle e_{j_1}\otimes\cdots\otimes e_{j_n}\right| \otimes\left|f_{i_1}\otimes\cdots\otimes f_{i_n}\right\rangle\left\langle f_{j_1}\otimes\cdots\otimes f_{j_n}\right|.$$

The first part of theorem 1 then follows immediately from lemma 4.

On the other hand, according to the spectral representation (4), one can identify the totality of output states $\{\mathrm{id}\otimes\Gamma_\theta(\tau^{ME})\}_\theta$ with the probability simplex $\mathcal{P}^{d^2-1}$ in the sense of quantum information geometry (cf, the appendix). The second part of theorem 1 now follows from the well-known fact that a classical statistical model has an efficient estimator if and only if the model is an exponential family and the parameter is mixture affine [7, theorem 3.12]. The proof is completed. □

3. ∇^e-autoparallel submodels

Theorem 1 implies that one can identify the family of generalized Pauli channels with the probability simplex $\mathcal{P}^{d^2-1}$ through the canonical embedding $\Gamma \mapsto \mathrm{id} \otimes \Gamma(\tau^{ME})$. This identification further allows us to treat submodels of generalized Pauli channels in a unified manner. According to theorems 2.5 and 3.12 in [7], each ∇^e-autoparallel submanifold $\mathcal{M}$ of $\mathcal{P}^{d^2-1}$ forms an exponential family and has an efficient estimator for the $\tilde{\nabla}^m$-affine coordinate system of $\mathcal{M}$, where $\tilde{\nabla}^m$ is the connection on $\mathcal{M}$ induced from the mixture connection ∇^m of $\mathcal{P}^{d^2-1}$. Therefore, each ∇^e-autoparallel submanifold of $\mathcal{P}^{d^2-1}$ corresponds to a statistically tractable submodel of generalized Pauli channels. We might as well call such a model a ∇^e-*autoparallel submodel*. Any other (smooth) submodel of generalized Pauli channels can be regarded as a curved exponential family [7], the treatment of which is a standard one in classical statistics. We summarize these observations by the following:

Theorem 5. *With the canonical embedding $\Gamma \mapsto \mathrm{id} \otimes \Gamma(\tau^{ME})$, a submodel of generalized Pauli channels admits an efficient estimator if and only if the submodel is ∇^e-autoparallel.*

An illustrative example of a ∇^e-autoparallel submodel is the family of depolarizing channels:

$$\Gamma_\eta(\tau) = \eta\tau + \frac{1-\eta}{d} I \qquad (\tau \in \mathcal{S}(\mathbb{C}^d))$$

which has a one-dimensional parameter $\eta \in \mathbb{R}$ that describes the magnitude of depolarization. The above equation can be transformed into the form (1), where

$$p_0 = \frac{1+(d^2-1)\eta}{d^2} \qquad p_k = \frac{1-\eta}{d^2} \qquad (1 \leqslant k \leqslant d^2-1).$$

The complete positivity condition for Γ_η, i.e., $p_k \geqslant 0$ for all k, is reduced to

$$-\frac{1}{d^2-1} \leqslant \eta \leqslant 1.$$

Geometrically, the family of depolarizing channels forms the straight line connecting the vertex $(1, 0, \ldots, 0)$ and the centre $(0, 1/(d^2-1), \ldots, 1/(d^2-1))$ of its opposite side in $\mathcal{P}^{d^2-1}$ (see figure 1). According to theorem 6, this line is not only a ∇^m-geodesic, but also a ∇^e-geodesic. As a consequence, the ∇^m-affine parameter η of the depolarizing channel has an efficient estimator. This completely solves the open problems posed in [1] and [4].

Let m and n be integers satisfying $2 \leqslant m \leqslant n$, and let $\{A^0, A^1, \ldots, A^{m-1}\}$ be a partition of the set $\{0, 1, \ldots, n-1\}$ into disjoint subsets. With each A^i, associate an n-dimensional probability vector $Q^i = \left(Q^i_0, \ldots, Q^i_{n-1}\right)$ having the support set A^i, that is, $Q^i_j > 0$ for $j \in A^i$ and $Q^i_j = 0$ otherwise. Define a congruent embedding $f : \mathcal{P}^{m-1} \to \mathcal{P}^{n-1} : (x_0, \ldots, x_{m-1}) \mapsto (X_0, \ldots, X_{n-1})$ by

$$X_j = \sum_{i=0}^{m-1} x_i Q^i_j.$$

An embedding of this type is sometimes referred to as a *Markov map* [12, 13]. For example, the above-mentioned straight line of depolarizing channels in $\mathcal{P}^{d^2-1}$ can be regarded as the image of a Markov map $f : \mathcal{P}^1 \to \mathcal{P}^{d^2-1}$ in which $A^0 = \{0\}$, $A^1 = \{1, \ldots, d^2-1\}$, and $Q^0 = (1, 0, \ldots, 0)$, $Q^1 = (0, 1/(d^2-1), \ldots, 1/(d^2-1))$.

Theorem 6. *The image $f(\mathcal{P}^{m-1})$ of a Markov map f is ∇^e-autoparallel in $\mathcal{P}^{n-1}$.*

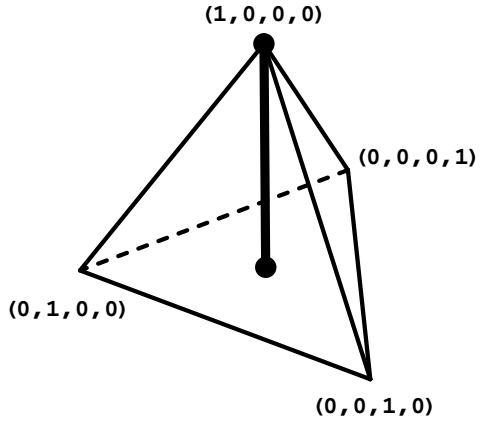

Figure 1. The family of depolarizing channels for $d = 2$, embedded in the probability simplex $\mathcal{P}^3$ of Pauli channels. It is not only a ∇^m-geodesic but also a ∇^e-geodesic of $\mathcal{P}^3$.

Proof. Observe that the image $f(\mathcal{P}^{m-1})$ is the interior of the convex hull of the points $Q^0, \ldots, Q^{m-1}$. Take arbitrary points

$$X = \sum_{i=0}^{m-1} x_i Q^i \qquad Y = \sum_{i=0}^{m-1} y_i Q^i$$

in $f(\mathcal{P}^{m-1})$. Since the supports of $Q^0, \ldots, Q^{m-1}$ are mutually disjoint, the ∇^e-geodesic connecting X and Y is written as

$$\frac{1}{Z(t)}((X_0)^t(Y_0)^{1-t}, \ldots, (X_{n-1})^t(Y_{n-1})^{1-t}) = \frac{1}{Z(t)}\sum_{i=0}^{m-1}(x_i)^t(y_i)^{1-t}Q^i \tag{6}$$

where $t(\in \mathbb{R})$ is the ∇^e-affine parameter and

$$Z(t) := \sum_{i=0}^{m-1}(x_i)^t(y_i)^{1-t}$$

is the normalization factor. Equation (6) shows that the ∇^e-geodesic connecting two arbitrary points in $f(\mathcal{P}^{m-1})$ runs through $f(\mathcal{P}^{m-1})$. In other words, $f(\mathcal{P}^{m-1})$ is totally ∇^e-geodesic, and hence is ∇^e-autoparallel [14, chapter VII, theorem 8.4] since the exponential connection ∇^e of $\mathcal{P}^{n-1}$ is torsion free. □

Let us return to the analysis of the depolarizing channel Γ_η, and demonstrate the simplest case $d = 2$ in more detail (see also [1]). Let $\tau^{ME} = |\psi\rangle\langle\psi|$, where

$$\psi = \frac{1}{\sqrt{2}}\left(\begin{bmatrix}1\\0\end{bmatrix}\otimes\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix}\otimes\begin{bmatrix}0\\1\end{bmatrix}\right).$$

The corresponding output states

$$\rho_\eta = \mathrm{id}\otimes\Gamma_\eta(\tau^{ME}) = \frac{1}{4}\begin{bmatrix}1+\eta & 0 & 0 & 2\eta\\ 0 & 1-\eta & 0 & 0\\ 0 & 0 & 1-\eta & 0\\ 2\eta & 0 & 0 & 1+\eta\end{bmatrix}$$

can be represented in the form of a (commutative) exponential family [7, theorem 7.6]:

$$\rho_\eta = \frac{I}{4}\exp[\beta(\eta)T - \gamma(\eta)I]$$

where

$$\beta(\eta) = \frac{3}{4}\log\frac{1+3\eta}{1-\eta} \qquad \gamma(\eta) = -\frac{1}{4}\log(1-\eta)^3(1+3\eta)$$

and

$$T = \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix}.$$

Thus the 'observable' T gives the efficient estimator for the parameter η $(= \gamma'(\eta)/\beta'(\eta))$. In fact, it is easy to check the unbiasedness $\mathrm{Tr}\,\rho_\eta T = \eta$ and the efficiency

$$\mathrm{Tr}\,\rho_\eta(T-\eta I)^2 = \frac{(1-\eta)(1+3\eta)}{3}\left(= \frac{1}{J_\eta(\tau^{ME})}\right)$$

for all $\eta \in [-1/3, 1]$. Needless to say, the optimal input and estimator for the nth extension $(\mathrm{id}\otimes\Gamma_\eta)^{\otimes n}$ are $(\tau^{ME})^{\otimes n}$ and $(1/n)\sum_{i=1}^n I^{\otimes(i-1)}\otimes T\otimes I^{\otimes(n-i)}$.

4. Discussions

We have presented the parameter estimation theory for a generalized Pauli channel $\Gamma_\theta : \mathcal{S}(\mathbb{C}^d) \to \mathcal{S}(\mathbb{C}^d)$. We have shown that, for each degree n of extension $(\mathrm{id}\otimes\Gamma_\theta)^{\otimes n}$, the SLD Fisher information matrix for the output states takes the maximum when the input is the n-tensor product of a maximally entangled state $\tau^{ME} \in \mathcal{S}(\mathbb{C}^d\otimes\mathbb{C}^d)$, and that there is an efficient estimator for the corresponding quantum Cramér–Rao inequality if and only if the parameter θ is ∇^m-affine in $\mathcal{P}^{d^2-1}$ (theorem 1). These results imply that the optimal strategy for estimating a generalized Pauli channel is of iid type: one need only repeat the optimal estimation procedure for the output state $\mathrm{id}\otimes\Gamma_\theta(\tau^{ME})$. The key observation to the proof of theorem 1, that the family $\{\mathrm{id}\otimes\Gamma_\theta(\tau^{ME})\}_\theta$ of output states can be identified with the classical statistical manifold $\mathcal{P}^{d^2-1}$, further allowed us to investigate submodels of generalized Pauli channels in a unified manner.

There is an alternative view for theorem 1 [15]. Given K unitary operators $\{V_k\}_{1\leqslant k\leqslant K}$ and a density operator τ on $\mathcal{H}$, let $\mathcal{A}$ be a K-dimensional commutative $*$-algebra, and define the map $\Lambda : \mathcal{A} \to \mathcal{B}(\mathcal{H})$ by

$$\Lambda : (a_k)_k \longmapsto \sum_{k=1}^K a_k(V_k\tau V_k^*).$$

Then Λ is completely positive and trace preserving, and when it is restricted to a parametric model $\{p(\theta)\}_\theta \subset \mathcal{P}^{K-1}$, we have $J(p(\theta)) \geqslant J(\Lambda(p(\theta)))$ because of the monotonicity [11], where $J(p(\theta))$ and $J(\Lambda(p(\theta)))$ are the classical Fisher information of $p(\theta)$ and the SLD Fisher information of $\Lambda(p(\theta))$. Obviously, this observation leads us to an alternative (in a sense, dual) proof of the first part of theorem 1, by showing that the upper bound $J(p(\theta))(=\hat{J}_\theta)$ is achievable. In fact, when $V_k = I\otimes\sigma_k$ $(0\leqslant k\leqslant d^2-1)$ and $\tau = \tau^{ME}$ on $\mathcal{H} = \mathbb{C}^d\otimes\mathbb{C}^d$, it holds that $J(p(\theta)) = J(\Lambda(p(\theta)))$ as we have seen in section 2.

Nevertheless, our approach allows us to proceed to a more detailed analysis: let us prove that there is no input state but a maximally entangled state τ^{ME} that is optimal in view of

theorem 1. For a pure state input $\tau^\alpha \in \mathcal{S}(\mathbb{C}^d \otimes \mathbb{C}^d)$, the output state $\rho_\theta^\alpha := \mathrm{id} \otimes \Gamma_\theta(\tau^\alpha)$ was decomposed into $\rho_\theta^\alpha = A^\alpha \rho_\theta A^\alpha$, where $A^\alpha := \sqrt{d}\sum_{i=1}^d \sqrt{\alpha_i}|e_i\rangle\langle e_i| \otimes I$ and $\rho_\theta := \rho_\theta^u (>0)$ with $u = (1/d, \ldots, 1/d)$. This shows that $\mathrm{rank}\,\rho_\theta^\alpha = \mathrm{rank}\,A^\alpha = rd$, where r is the number of nonzero components in α. Therefore, in order for the family $\{\rho_\theta^\alpha\}_\theta$ to be (d^2-1)-dimensional and commutative (to ensure that the SLD and the RLD Fisher information are identical on $\mathrm{supp}\,\rho_\theta^\alpha = \mathrm{supp}\,A^\alpha$), r must be equal to d, namely $\alpha > 0$. In this case, the family $\{\rho_\theta^\alpha\}_\theta$ is commutative if and only if $\alpha = u$. The 'if' part follows from lemma 2. To prove the 'only if' part, let

$$\rho_\theta = \sum_{k=0}^{d^2-1} p_k(\theta)|g_k\rangle\langle g_k|$$

be the simultaneous spectral decomposition of the commutative family $\{\rho_\theta\}_\theta$, where $g_k := (1/\sqrt{d})\sum_{i=1}^d e_i \otimes (\sigma_k f_i)$. Then the assumption, that $[\rho_\theta^\alpha, \rho_{\theta'}^\alpha] = 0$ for all θ, θ', is equivalent to $[(A^\alpha)^2, \rho_\theta] = 0$ for all θ (to see the necessity, just let $\rho_{\theta'} = (I/d)^{\otimes 2}$), and is further equivalent to $[(A^\alpha)^2, |g_k\rangle\langle g_k|] = 0$ for all k. On the other hand, by a direct calculation

$$[(A^\alpha)^2, |g_k\rangle\langle g_k|] = \sum_{i,j=1}^d (\alpha_i - \alpha_j)|e_i\rangle\langle e_j| \otimes |\sigma_k f_i\rangle\langle \sigma_k f_j|.$$

We thus have $\alpha_i = \alpha_j$ for all i, j, which proves the claim. These observations sharpen theorem 1, in that the optimal input is unique up to the Schmidt bases.

It should be noted that for models that lie in the boundary set of $\mathcal{P}^{d^2-1}$ (which have been excluded in our analysis), non-maximally entangled inputs might perform as well as a maximally entangled one. For example, consider the phase damping channel $\Gamma_\eta : \mathcal{S}(\mathbb{C}^2) \to \mathcal{S}(\mathbb{C}^2)$ defined by

$$\Gamma_\eta(\tau) = \eta\tau + (1-\eta)(\sigma_3\tau\sigma_3^*) \qquad (0 \leqslant \eta \leqslant 1).$$

Geometrically, this model corresponds to the side connecting the vertices $(1, 0, 0, 0)$ and $(0, 0, 0, 1)$ in $\mathcal{P}^3$ (see figure 1), and η is a ∇^m-affine parameter. Then by a continuity argument, one can deduce from theorems 1 and 5 that a maximally entangled state τ^{ME} is an optimal input, and η has an efficient estimator. Now let $\tau^\alpha = |\phi_0^\alpha\rangle\langle\phi_0^\alpha|$, where

$$\phi_0^\alpha = \sqrt{1-\alpha}\begin{bmatrix}1\\0\end{bmatrix} \otimes \begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix} + \sqrt{\alpha}\begin{bmatrix}0\\1\end{bmatrix} \otimes \begin{bmatrix}-1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}.$$

Then the corresponding output state $\rho_\eta^\alpha = \mathrm{id} \otimes \Gamma_\eta(\tau^\alpha)$ becomes

$$\rho_\eta^\alpha = \eta|\phi_0^\alpha\rangle\langle\phi_0^\alpha| + (1-\eta)|\phi_1^\alpha\rangle\langle\phi_1^\alpha| \tag{7}$$

where

$$\phi_1^\alpha = \sqrt{1-\alpha}\begin{bmatrix}1\\0\end{bmatrix} \otimes \begin{bmatrix}-1\sqrt{2}\\1\sqrt{2}\end{bmatrix} + \sqrt{\alpha}\begin{bmatrix}0\\1\end{bmatrix} \otimes \begin{bmatrix}1\sqrt{2}\\1\sqrt{2}\end{bmatrix}.$$

Since $\langle\phi_0^\alpha|\phi_1^\alpha\rangle = 0$, equation (7) implies that for all $\alpha \in [0, 1]$, the family $\{\rho_\eta^\alpha\}_\eta$ is commutative, and is isomorphic to a coin flipping in which 'heads' occur with probability η. This shows that the SLD Fisher information $J_\eta(\tau^\alpha)$ is independent of α, and is identical to the classical Fisher information $1/\eta(1-\eta)$. As a consequence, for estimating a phase damping channel, the inputs $\tau^\alpha = |\phi_0^\alpha\rangle\langle\phi_0^\alpha|$ perform equally well for all degrees α of entanglement. A similar argument applies to the bit flip channel. Such an anomaly is due to the degeneracy of the output states $\rho_\eta^u = \mathrm{id} \otimes \Gamma_\eta(\tau^{ME})$, and does not contradict the argument presented in the preceding paragraph (cf, [16–18]).

Acknowledgment

We thank M Hayashi for a valuable comment.

Appendix. Quantum information geometry

This appendix provides a brief account of quantum information geometry based on the SLD. Let $\mathcal{S}$ be the totality of faithful quantum states on a D-dimensional Hilbert space $\mathcal{H}$. The set $\mathcal{S}$ is naturally regarded as a (D^2-1)-dimensional differentiable manifold, and its dualistic geometrical structure is introduced as follows. We first define a Riemannian metric by

$$g(X, Y) := \tfrac{1}{2}\operatorname{Tr}\rho(L_X L_Y + L_Y L_X) = \operatorname{Tr}(X\rho)L_Y$$

where $X, Y \in T_\rho\mathcal{S}$, and L_X, L_Y are the corresponding SLDs, i.e., the Hermitian operators satisfying

$$X\rho = \tfrac{1}{2}(\rho L_X + L_X\rho).$$

The metric g is called the *SLD Fisher* metric. We next introduce a pair of affine connections. One is defined by

$$\big(\nabla^m_X Y\big)\rho := X(Y\rho)$$

and is called the *mixture* connection. The other is defined by

$$\big(\nabla^e_X Y\big)\rho := \tfrac{1}{2}\{\rho(XL_Y - \operatorname{Tr}\rho(XL_Y)) + (XL_Y - \operatorname{Tr}\rho(XL_Y))\rho\}$$

and is called the *exponential* connection. These connections are mutually dual with respect to the SLD Fisher metric, in that

$$Xg(Y, Z) = g\big(\nabla^m_X Y, Z\big) + g\big(Y, \nabla^e_X Z\big).$$

A coordinate system $\xi = (\xi^i)_{1\leqslant i\leqslant D^2-1}$ of $\mathcal{S}$ is called *affine* with respect to a connection ∇ of $\mathcal{S}$ if $\nabla_{\partial_i}\partial_j = 0$ for all i, j, where $\partial_i = \partial/\partial\xi^i$. For example, the components of density matrices $\rho(\in\mathcal{S})$, with one diagonal entry removed (since $\operatorname{Tr}\rho = 1$), form a ∇^m-affine coordinate system of $\mathcal{S}$. On the other hand, $\mathcal{S}$ does not have a ∇^e-affine coordinate system, since ∇^e-torsion does not vanish because of the noncommutativity of operators.

A submanifold $\mathcal{M}$ of $\mathcal{S}$ is called *autoparallel* with respect to a connection ∇ of $\mathcal{S}$ if $\nabla_X Y \in T_\rho\mathcal{M}$ for all $\rho \in \mathcal{M}$ and $X, Y \in T_\rho\mathcal{M}$. In particular, a one-dimensional ∇-autoparallel submanifold is called a ∇-*geodesic*. When $\mathcal{M}$ is ∇-autoparallel in $\mathcal{S}$, $\mathcal{M}$ has a vanishing embedding curvature with respect to ∇, and one can regard ∇ as a connection of $\mathcal{M}$, just by restricting ∇ onto $\mathcal{M}$. For example, a maximal commutative subset $\mathcal{P}$ of $\mathcal{S}$ is autoparallel with respect to both ∇^m and ∇^e, so that one can naturally induce a dualistic structure on $\mathcal{P}$ from that of $\mathcal{S}$. In fact, the geometrical structure thus induced on $\mathcal{P}$ is isomorphic to that of the classical probability simplex $\mathcal{P}^{D-1}$. Such an isomorphism is usefully exploited throughout the paper. For more information, see [7]. A generalization to manifolds of non-faithful (i.e., degenerate) quantum states is discussed in [18].

References

[1] Fujiwara A 2001 Quantum channel identification problem *Phys. Rev.* A **63** 042304
[2] Acín A, Jané E and Vidal G 2001 Optimal estimation of quantum dynamics *Phys. Rev.* A **64** 050302(R)
[3] Fujiwara A 2002 Estimation of $SU(2)$ operation and dense coding: an information geometric approach *Phys. Rev.* A **65** 012316
[4] Sasaki M, Ban M and Barnett S M 2002 Optimal parameter estimation of a depolarizing channel *Phys. Rev.* A **66** 022308

[5] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
[6] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
[7] Amari S and Nagaoka H 2000 *Methods of Information Geometry (Transl. Math. Monographs vol 191)* (Providence, RI: American Mathematical Society)
[8] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[9] Cirone M A, Delgado A, Fischer D G, Freyberger M, Mack H and Mussinger M 2001 Estimation of quantum channels with finite resources *Preprint* quant-ph/0108037
[10] Fivel D I 1995 Remarkable phase oscillations appearing in the lattice dynamics of Einstein–Podolsky–Rosen states *Phys. Rev. Lett.* **74** 835–8
[11] Petz D 1996 Monotone metrics on matrix spaces *Linear Algebr. Appl.* **244** 81–96
[12] Čencov N N 1981 *Statistical Decision Rules and Optimal Inference (Transl. Math. Monographs vol 53)* (Providence, RI: American Mathematical Society)
[13] Campbell L L 1986 An extended Čencov characterization of the information metric *Proc. Am. Math. Soc.* **98** 135–41
[14] Kobayashi S and Nomizu K 1969 *Foundations of Differential Geometry* vol II (New York: Interscience)
[15] Hayashi M Private communication
[16] Fujiwara A and Nagaoka H 1995 Quantum Fisher metric and estimation for pure state models *Phys. Lett.* A **201** 119–24
[17] Fujiwara A and Nagaoka H 1999 An estimation theoretical characterization of coherent states *J. Math. Phys.* **40** 4227–39
[18] Fujiwara A 1999 Geometry of quantum information systems *Geometry in Present Day Sciences* ed O E Barndorff-Nielsen and E B V Jensen (Singapore: World Scientific) pp 35–48