entropy

MDPI

Article

Information Geometry of Randomized Quantum State Tomography

Akio Fujiwara [1,*] and Koichi Yamagata [2]

1 Department of Mathematics, Osaka University, Toyonaka, Osaka 560-0043, Japan
2 Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan; koichi.yamagata@uec.ac.jp
* Correspondence: fujiwara@math.sci.osaka-u.ac.jp

Received: 29 June 2018; Accepted: 13 August 2018; Published: 16 August 2018

Abstract: Suppose that a d-dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$ admits a full set of mutually unbiased bases $\left\{|1^{(a)}\rangle, \ldots, |d^{(a)}\rangle\right\}$, where $a = 1, \ldots, d+1$. A randomized quantum state tomography is a scheme for estimating an unknown quantum state on $\mathcal{H}$ through iterative applications of measurements $M^{(a)} = \left\{|1^{(a)}\rangle\langle 1^{(a)}|, \ldots, |d^{(a)}\rangle\langle d^{(a)}|\right\}$ for $a = 1, \ldots, d+1$, where the numbers of applications of these measurements are random variables. We show that the space of the resulting probability distributions enjoys a mutually orthogonal dualistic foliation structure, which provides us with a simple geometrical insight into the maximum likelihood method for the quantum state tomography.

Keywords: quantum state tomography; mutually unbiased bases; information geometry; dualistic foliation; mixed coordinate system

1. Introduction

Quantum state tomography is a method of estimating an unknown quantum state represented on some Hilbert space $\mathcal{H}$, consisting of a fixed set of measurements that provides sufficient information about the unknown quantum state, as well as a data processing that maps each measurement outcome into the quantum state space $\mathcal{S}(\mathcal{H})$ on $\mathcal{H}$ [1]. A set of measurements that fulfils this requirement is sometimes called a measurement basis. For mathematical simplicity, we restrict ourselves to Hilbert spaces of finite dimensions.

To elucidate our motivation, let us treat the simplest case when $\mathcal{H} \simeq \mathbb{C}^2$. It is well known that there is a one-to-one affine correspondence between the qubit state space

$$\mathcal{S}(\mathbb{C}^2) := \{\rho \in \mathbb{C}^{2\times 2} \mid \rho \geq 0,\ \mathrm{Tr}\,\rho = 1\}$$

and the unit ball (called the Bloch ball)

$$B := \left\{x = (x_1, x_2, x_3) \in \mathbb{R}^3 \;\middle|\; \|x\|^2 := (x_1)^2 + (x_2)^2 + (x_3)^2 \leq 1\right\}.$$

In fact, the correspondence is explicitly given by the Stokes parametrization

$$x \longmapsto \rho_x = \frac{1}{2}(I + x_1\sigma_1 + x_2\sigma_2 + x_3\sigma_3),$$

where σ_1, σ_2, and σ_3 are the standard Pauli matrices. Since $E_{\rho_x}[\sigma_i] := \mathrm{Tr}\,\rho_x\sigma_i = x_i$ for $i \in \{1,2,3\}$, the set $\sigma = (\sigma_1, \sigma_2, \sigma_3)$ of observables is regarded as an unbiased estimator [2–4] for the Stokes parameter $x = (x_1, x_2, x_3)$. This is the basic idea behind the standard qubit state tomography, which runs as follows: suppose that, among N independent experiments, the ith Pauli matrix σ_i was measured $N/3$ times, and outcomes $+1$ (spin-up) and -1 (spin-down) were obtained n_i^+ and n_i^- times, respectively. Then a naive estimate for the true value of the parameter $x = (x_1, x_2, x_3)$ is

$$\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3) := \left(\frac{n_1^+ - n_1^-}{N/3}, \frac{n_2^+ - n_2^-}{N/3}, \frac{n_3^+ - n_3^-}{N/3} \right).$$

When the estimate $\hat{x} \in [-1,1]^3$ falls outside the Bloch ball B, it needs to be corrected so that the new estimate lies in the Bloch ball B. The maximum likelihood method is a canonical one to obtain a corrected estimate [2,5–10]. From the point of view of information geometry [11–13], the maximum likelihood estimate (MLE) is the orthogonal projection from the temporary estimate $\hat{x}$ onto the Bloch ball B with respect to the standard Fisher metric along the $\nabla^{(m)}$-geodesic [14], (cf., Appendix A).

Now let us deal with a slightly generalized situation: suppose that the ith Pauli matrix σ_i was measured N_i times and outcomes $+1$ and -1 were obtained n_i^+ and n_i^- times, respectively, where $\{N_i\}_{i=1,2,3}$ were random variables. Such a situation arises in an actual experiment due to unexpected particle loss [15]. We shall call such a generalized estimation scheme a *randomized state tomography*. A naive estimate in this case is the following:

$$\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3) := \left(\frac{n_1^+ - n_1^-}{N_1}, \frac{n_2^+ - n_2^-}{N_2}, \frac{n_3^+ - n_3^-}{N_3} \right).$$

One may invoke the maximum likelihood method when $\hat{x}$ falls outside the Bloch ball. It is then interesting to ask if there is also a useful geometrical picture for the MLE even when the numbers N_i of measurements are random variables.

The above mentioned problem is naturally extended to quantum state tomography on an arbitrary Hilbert space that admits a full set of mutually unbiased bases [16,17]. In a d-dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$, k orthonormal bases

$$\left\{ |\alpha^{(1)}\rangle \right\}_{\alpha \in \{1,\dots,d\}}, \left\{ |\beta^{(2)}\rangle \right\}_{\beta \in \{1,\dots,d\}}, \dots, \left\{ |\gamma^{(k)}\rangle \right\}_{\gamma \in \{1,\dots,d\}}$$

are called *mutually unbiased* if they satisfy

$$\left| \left\langle \alpha^{(a)} \,\middle|\, \beta^{(b)} \right\rangle \right|^2 = \frac{1}{d}$$

for all $a, b \in \{1, \dots, k\}$ with $a \neq b$, and $\alpha, \beta \in \{1, \dots, d\}$. It is known that the number k of mutually unbiased bases (MUBs) is at most $d+1$ [18]. If there are $d+1$ MUBs, the Hilbert space $\mathcal{H}$ is said to admit a full set of MUBs. For example, when the dimension d of $\mathcal{H}$ is a power of a prime, $\mathcal{H}$ admits a full set of MUBs [19]. Whether or not any Hilbert space admits a full set of MUBs is an open question [16].

In what follows, unless otherwise stated, we assume that the Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$ under consideration admits a full set of MUBs. As demonstrated in Appendix B (cf., [17,20]), each density operator $\rho \in \mathcal{S}(\mathcal{H})$ can be uniquely represented as

$$\rho = \rho(\xi) := \sum_{a=1}^{d+1} \left\{ \sum_{\alpha=1}^{d-1} \xi_\alpha^{(a)} M_\alpha^{(a)} + \left(1 - \sum_{\beta=1}^{d-1} \xi_\beta^{(a)} \right) M_d^{(a)} \right\} - I, \tag{1}$$

where

$$M^{(a)} := \left\{ M_1^{(a)}, \dots, M_d^{(a)} \right\} = \left\{ |1^{(a)}\rangle \langle 1^{(a)}|, \dots, |d^{(a)}\rangle \langle d^{(a)}| \right\}$$

is the projection-valued measure (PVM) associated with the ath orthogonal basis in the MUBs, and

$$\xi := \left(\xi_\alpha^{(a)}\right)_{(a,\alpha)\in\{1,\dots,d+1\}\times\{1,\dots,d-1\}}$$

is a (d^2-1)-dimensional real parameter that is chosen so that $\rho(\xi)\geq 0$. A simple calculation shows that, if the ath measurement $M^{(a)}$ is applied to the state $\rho(\xi)$, one obtains each outcome $\alpha\in\{1,\dots,d\}$ with probability

$$p_\alpha^{(a)} = \mathrm{Tr}\,\rho(\xi)M_\alpha^{(a)} = \begin{cases} \xi_\alpha^{(a)}, & \text{for } \alpha = 1,\dots,d-1, \\ 1-\sum_{\beta=1}^{d-1}\xi_\beta^{(a)}, & \text{for } \alpha = d. \end{cases} \tag{2}$$

This implies that the parametrization $\xi\mapsto\rho(\xi)$ establishes an affine isomorphism between the quantum state space

$$\mathcal{S}(\mathbb{C}^d) := \left\{\rho\in\mathbb{C}^{d\times d}\,\middle|\,\rho\geq 0,\ \mathrm{Tr}\,\rho = 1\right\}$$

and the convex set

$$B := \left\{\xi\in\mathbb{R}^{d^2-1}\,\middle|\,\rho(\xi)\geq 0\right\}.$$

Incidentally, the Stokes parametrization $x\mapsto\rho_x$ for the qubit state space $\mathcal{S}(\mathbb{C}^2)$ is regarded as a special case of the above parametrization $\xi\mapsto\rho(\xi)$ for $\mathcal{S}(\mathbb{C}^d)$. In fact, the eigenvectors of the Pauli matrices σ_1, σ_2, σ_3 form a full set of MUBs on $\mathbb{C}^2$, and the Stokes parametrization $x=(x_1,x_2,x_3)$ is related to the above parametrization $\xi=(\xi_1^{(1)},\xi_1^{(2)},\xi_1^{(3)})$ as

$$\xi_1^{(a)} = \frac{x_a+1}{2}, \qquad (a=1,2,3).$$

Now that a standard affine parametrization $\xi\mapsto\rho(\xi)$ has been established on an arbitrary Hilbert space $\mathcal{H}\simeq\mathbb{C}^d$ that admits a full set of MUBs, the scheme of randomized state tomography is naturally extended to $\mathcal{H}$ as follows. Suppose that the ath measurement $M^{(a)}$ was applied $N^{(a)}$ times and the outcome $\alpha\in\{1,\dots,d\}$ was obtained $n_\alpha^{(a)}$ times, where $\left\{N^{(a)}\right\}_{a=1,\dots,d+1}$ were random variables. Then, due to (2), a naive estimate for the parameter $\xi_\alpha^{(a)}$ is

$$\hat{\xi}_\alpha^{(a)} = \frac{n_\alpha^{(a)}}{N^{(a)}}.$$

When the estimate $\hat{\xi} := \left(\hat{\xi}_\alpha^{(a)}\right)\in[0,1]^{d^2-1}$ falls outside the parameter space B, one may invoke the maximum likelihood method to obtain a corrected estimate.

The objective of the present paper is to clarify that the $\nabla^{(m)}$-projection interpretation for the MLE is still valid for the randomized state tomography by changing the standard Fisher metric into a deformed one depending on the realization of the random variables $N^{(a)}$, which might as well be called a randomized Fisher metric. Such a novel geometrical picture will provide important insights into the quantum metrology.

The paper is organized as follows. In Section 2, we first introduce a statistical model on an extended sample space Ω that represents the randomized state tomography. We then clarify that the probability simplex $\mathcal{P}(\Omega)$ is decomposed into mutually orthogonal dualistic foliation by means of certain $\nabla^{(m)}$- and $\nabla^{(e)}$-autoparallel submanifolds. In Section 3, we give a statistical interpretation of the above-mentioned dualistic foliation structure. In particular, we point out that the MLE is the $\nabla^{(m)}$-projection with respect to a deformed Fisher metric that depends on the realization of the random variables $N^{(a)}$. These results are demonstrated by several illustrative examples in Section 4. Finally, some concluding remarks are presented in Section 5. For the reader's convenience, some background information is provided in Appendices A and B, including information geometry of the MLE and affine parametrization of a quantum state space $\mathcal{S}(\mathcal{H})$.

2. Geometry of Randomized State Tomography

We identify the randomized state tomography on $\mathcal{H} \simeq \mathbb{C}^d$ with the following scheme [21]: at each step of the measurement, one chooses a PVM $M^{(a)}$ at random with probability $s^{(a)}$, $(a = 1, \ldots, d+1)$, and applies the chosen PVM to yield an outcome $\alpha \in \{1, \ldots, d\}$. The sample space Ω for this statistical picture is

$$\Omega = \{(a, \alpha) \,|\, a \in \{1, \ldots, d+1\},\, \alpha \in \{1, \ldots, d\}\}.$$

Suppose that the unknown state ρ is specified by the coordinate $\xi \in B$ as (1). Then the corresponding probability distribution on Ω is represented by the $d(d+1)$-dimensional probability vector

$$p_{(s,\xi)} := \left(s^{(1)} \left(\xi_1^{(1)}, \ldots, \xi_{d-1}^{(1)}, 1 - \sum_{\alpha=1}^{d-1} \xi_\alpha^{(1)} \right), \ldots, s^{(d)} \left(\xi_1^{(d)}, \ldots, \xi_{d-1}^{(d)}, 1 - \sum_{\alpha=1}^{d-1} \xi_\alpha^{(d)} \right), \right.$$
$$\left. \left(1 - \sum_{a=1}^{d} s^{(a)} \right) \left(\xi_1^{(d+1)}, \ldots, \xi_{d-1}^{(d+1)}, 1 - \sum_{\alpha=1}^{d-1} \xi_\alpha^{(d+1)} \right) \right)$$

where the parameter $s := (s^{(1)}, \ldots, s^{(d)})$ belongs to the domain

$$D := \left\{ s \in \mathbb{R}^d \,\middle|\, s^{(a)} > 0 \text{ for } a \in \{1, \ldots, d\}, \text{ and } \sum_{a=1}^{d} s^{(a)} < 1 \right\}.$$

Note that the family

$$\left\{ p_{(s,\xi)} \middle| s \in D,\, \xi \in \Xi \right\}$$

with

$$\Xi := \left\{ \xi \in \mathbb{R}^{d^2-1} \,\middle|\, \xi_\alpha^{(a)} > 0 \text{ for } (a, \alpha) \in \{1, \ldots, d+1\} \times \{1, \ldots, d-1\}, \right.$$
$$\left. \text{and } \sum_{\alpha=1}^{d-1} \xi_\alpha^{(a)} < 1 \text{ for } a \in \{1, \ldots, d+1\} \right\}$$

forms a $(d^2 + d - 1)$-dimensional open probability simplex $\mathcal{P}(\Omega)$, and the parameters (s, ξ) form a coordinate system of $\mathcal{P}(\Omega)$. Since we are only interested in estimating the parameter $\xi \in \Xi$, the remaining parameter $s \in D$ is understood as a set of nuisance parameters [2,12]. In what follows, we regard $\mathcal{P}(\Omega)$ as a statistical manifold endowed with the standard dualistic structure $\left(g, \nabla^{(e)}, \nabla^{(m)}\right)$, where g is the Fisher metric, and $\nabla^{(e)}$ and $\nabla^{(m)}$ are the exponential and mixture connections [12].

Let us consider the following submanifolds of $\mathcal{P}(\Omega)$:

$$M(s) := \left\{ p_{(s,\xi)} \mid \xi \in \Xi \right\}$$

for each $s \in D$, and

$$E(\xi) := \left\{ p_{(s,\xi)} \mid s \in D \right\}$$

for each $\xi \in \Xi$. Since $M(s)$ and $E(\xi)$ are convex subsets of $\mathcal{P}(\Omega)$, they are both $\nabla^{(m)}$-autoparallel. In addition, we have the following.

Proposition 1. *For each $\xi \in \Xi$, the submanifold $E(\xi)$ is $\nabla^{(e)}$-autoparallel. Furthermore, for each $s \in D$ and $\xi \in \Xi$, the submanifolds $M(s)$ and $E(\xi)$ are mutually orthogonal with respect to the Fisher metric g.*

Proof. Let us change the coordinate system (s, ξ) into $(\eta_{\langle a\rangle}, \eta_{\langle b,\alpha\rangle})$, where

$$\eta_{\langle a\rangle} := s^{(a)}$$

for $a \in \{1, \ldots, d\}$, and

$$\eta_{\langle b,\alpha\rangle} := s^{(b)}\xi_{\alpha}^{(b)}$$

for $(b, \alpha) \in \{1, \ldots, d+1\} \times \{1, \ldots, d-1\}$. With this coordinate transformation, the probability vector $p_{(s,\xi)}$ is rewritten as

$$p_\eta = \bigoplus_{a=1}^{d+1} \left(\eta_{\langle a,1\rangle}, \ldots, \eta_{\langle a,d-1\rangle}, \eta_{\langle a\rangle} - \sum_{\alpha=1}^{d-1} \eta_{\langle a,\alpha\rangle} \right). \tag{3}$$

Here, $\eta_{\langle d+1\rangle}$ is a function of $\{\eta_{\langle a\rangle}\}_{a\in\{1,\ldots,d\}}$ defined by

$$\eta_{\langle d+1\rangle} := 1 - \sum_{a=1}^{d} \eta_{\langle a\rangle},$$

and is not a component of the coordinate system $\eta := (\eta_{\langle a\rangle}, \eta_{\langle b,\alpha\rangle})$. We see from the representation (3) that the coordinate system η is $\nabla^{(m)}$-affine. The potential function for η is given by the negative entropy

$$\begin{aligned} \varphi(\eta) &:= \sum_{\omega\in\Omega} p_\eta(\omega) \log p_\eta(\omega) \\ &= \sum_{a=1}^{d+1} \left\{ \sum_{\alpha=1}^{d-1} \eta_{\langle a,\alpha\rangle} \log \eta_{\langle a,\alpha\rangle} + \left(\eta_{\langle a\rangle} - \sum_{\beta=1}^{d-1} \eta_{\langle a,\beta\rangle} \right) \log \left(\eta_{\langle a\rangle} - \sum_{\beta=1}^{d-1} \eta_{\langle a,\beta\rangle} \right) \right\} \end{aligned}$$

and the dual $\nabla^{(e)}$-affine coordinate system θ is given by

$$\theta^{\langle a\rangle} = \frac{\partial\varphi}{\partial\eta_{\langle a\rangle}} = \log \frac{s^{(a)}}{\left(1 - \sum_{b=1}^{d} s^{(b)}\right)} + \log \frac{\left(1 - \sum_{\beta=1}^{d-1} \xi_\beta^{(a)}\right)}{\left(1 - \sum_{\beta=1}^{d-1} \xi_\beta^{(d+1)}\right)}$$

for $a \in \{1, \ldots, d\}$, and

$$\theta^{\langle b,\alpha\rangle} = \frac{\partial\varphi}{\partial\eta_{\langle b,\alpha\rangle}} = \log \frac{\xi_\alpha^{(b)}}{\left(1 - \sum_{\beta=1}^{d-1} \xi_\beta^{(b)}\right)}$$

for $(b, \alpha) \in \{1, \ldots, d+1\} \times \{1, \ldots, d-1\}$. Thus, fixing ξ is equivalent to fixing the coordinates $\left(\theta^{\langle b,\alpha\rangle}\right)_{(b,\alpha)\in\{1,\ldots,d+1\}\times\{1,\ldots,d-1\}}$, and the submanifold $E(\xi)$ is generated by changing the remaining parameters $\left(\theta^{\langle a\rangle}\right)_{a\in\{1,\ldots,d\}}$. This implies that $E(\xi)$ is $\nabla^{(e)}$-autoparallel, proving the first part of the claim.

To prove the second part, let us introduce a mixed coordinate system [11]

$$\left(\eta_{\langle a\rangle}; \theta^{\langle b,\alpha\rangle}\right)_{a\in\{1,\ldots,d\},\,(b,\alpha)\in\{1,\ldots,d+1\}\times\{1,\ldots,d-1\}}$$

of $\mathcal{P}(\Omega)$. Since $\eta_{\langle a\rangle} = s^{(a)}$, the submanifold $M(s)$ is rewritten as

$$M(s) = \left\{ p_{(s,\xi)} \,\middle|\, \left(\eta_{\langle a\rangle}\right)_{a\in\{1,\ldots,d\}} \text{ are fixed and } \left(\theta^{\langle b,\alpha\rangle}\right)_{(b,\alpha)\in\{1,\ldots,d+1\}\times\{1,\ldots,d-1\}} \text{ are arbitrary} \right\}.$$

On the other hand, the submanifold $E(\xi)$ is rewritten as

$$E(\xi) = \left\{ p_{(s,\xi)} \,\middle|\, \left(\theta^{\langle b,\alpha\rangle}\right)_{(b,\alpha)\in\{1,\ldots,d+1\}\times\{1,\ldots,d-1\}} \text{ are fixed and } \left(\eta_{\langle a\rangle}\right)_{a\in\{1,\ldots,d\}} \text{ are arbitrary} \right\}.$$

Thus, the orthogonality of $M(s)$ and $E(\xi)$ is an immediate consequence of the orthogonality of the dual affine coordinate systems θ and η with respect to the Fisher metric g. $\square$

Proposition 1 implies that the manifold $\mathcal{P}(\Omega)$ is decomposed into mutually orthogonal dualistic foliation based on the submanifolds $M(s)$ and $E(\xi)$, as illustrated in Figure 1. We shall exploit this geometrical structure in the next section.

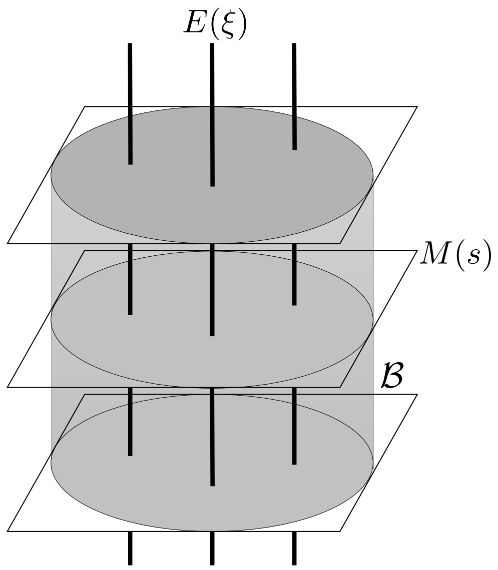

Figure 1. Mutually orthogonal dualistic foliation of $\mathcal{P}(\Omega)$ based on $M(s)$ and $E(\xi)$. Each section $M(s)$ is affinely isomorphic to the parameter space Ξ. The greyish cylindrical area indicates the subset $\mathcal{B} = \{p_{(s,\xi)} | s \in D,\, \xi \in B\}$ of $\mathcal{P}(\Omega)$. In particular, for each $s \in D$, the intersection $M(s) \cap \mathcal{B}$ is affinely isomorphic to the physical domain B that corresponds to the state space $\mathcal{S}(\mathcal{H})$.

3. Estimation of the Parameter ξ

Let us proceed to the problem of estimating the unknown parameter ξ using the randomized tomography. Suppose that, among N independent repetitions of experiments, the ath measurement $M^{(a)}$ was applied $N^{(a)}$ times and outcomes $\alpha \in \{1,\ldots,d\}$ were obtained $n_\alpha^{(a)}$ times. Then temporary estimates $(\hat{s},\hat{\xi})$ for the parameters (s,ξ) are given by

$$\hat{s}^{(a)} := \frac{N^{(a)}}{N}$$

for $a \in \{1,\ldots,d\}$, and

$$\hat{\xi}_\beta^{(b)} := \frac{n_\beta^{(b)}}{N^{(b)}}$$

for $(b,\beta) \in \{1,\ldots,d+1\} \times \{1,\ldots,d-1\}$. If $\hat{\xi}$ has fallen outside the physical domain B, one may seek a corrected estimate by the maximum likelihood method. Observe that, due to (2), the empirical distribution $\hat{q}_N \in \mathcal{P}(\Omega)$ is represented as

$$\hat{q}_N = p_{(\hat{s},\hat{\xi})}. \tag{4}$$

On the other hand, the physical domain B in the parameter space Ξ corresponds to the subset

$$\mathcal{B} := \{p_{(s,\xi)} \mid s \in D,\ \xi \in B\}$$

of $\mathcal{P}(\Omega)$, (see Figure 1). The MLE p^* in $\mathcal{P}(\Omega)$ is then given by

$$p^* = \arg\min_{p \in \mathcal{B}} D(\hat{q}_N \| p), \tag{5}$$

where $D(\cdot \| \cdot)$ is the Kullback-Leibler divergence (cf., Appendix A). A crucial observation is the following.

Proposition 2. *The minimum in (5) is achieved on $M(\hat{s}) \cap \mathcal{B}$.*

Proof. Let us take a point $p_{(s,\xi)} \in \mathcal{B}$ arbitrarily. It then follows from the mutually orthogonal dualistic foliation of $\mathcal{P}(\Omega)$ established in Proposition 1 that

$$\begin{aligned} D(\hat{q}_N \| p_{(s,\xi)}) &= D(p_{(\hat{s},\hat{\xi})} \| p_{(s,\xi)}) \\ &= D(p_{(\hat{s},\hat{\xi})} \| p_{(\hat{s},\xi)}) + D(p_{(\hat{s},\xi)} \| p_{(s,\xi)}) \\ &\geq D(p_{(\hat{s},\hat{\xi})} \| p_{(\hat{s},\xi)}). \end{aligned}$$

In the second equality, the generalized Pythagorean theorem was used. Consequently,

$$\min_{\xi \in B} D(p_{(\hat{s},\hat{\xi})} \| p_{(s,\xi)}) \geq \min_{\xi \in B} D(p_{(\hat{s},\hat{\xi})} \| p_{(\hat{s},\xi)})$$

for all $s \in D$, and the right-hand side is achieved if and only if $s = \hat{s}$. □

The geometrical implication of Proposition 2 is illustrated in Figure 2. The MLE $p^* = p_{(\hat{s},\xi^*)}$ is the $\nabla^{(m)}$-projection from the empirical distribution $p_{(\hat{s},\hat{\xi})}$ to $\mathcal{B}$, and is on the section $M(\hat{s})$ specified by the temporary estimate $\hat{s}$.

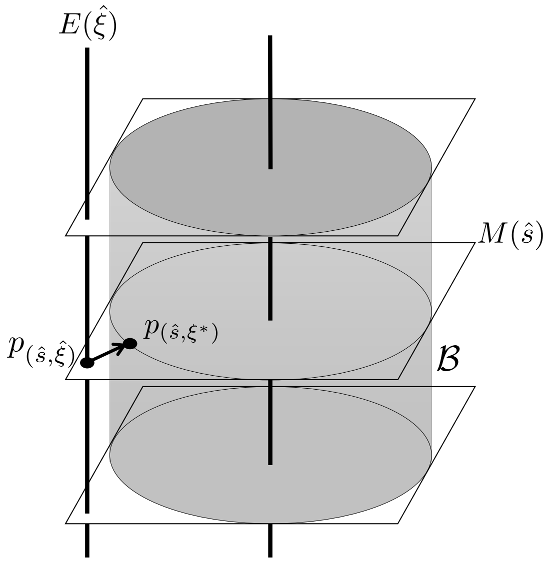

Figure 2. The maximum likelihood method in the framework of randomized tomography. Given a temporary estimate $(\hat{s}, \hat{\xi})$ with $\hat{\xi} \notin B$, we can restrict ourselves to the section $M(\hat{s})$ as the search space for the MLE p^*, and $p^* = p_{(\hat{s},\xi^*)}$ is the $\nabla^{(m)}$-projection from the empirical distribution $p_{(\hat{s},\hat{\xi})}$ to $\mathcal{B}$ on the section $M(\hat{s})$.

Now we arrive at a geometrical picture behind the parameter estimation based on randomized state tomography. Suppose we are given a temporary estimate $(\hat{s}, \hat{\xi})$ with $\hat{\xi} \notin B$. Due to Proposition 2, we can restrict ourselves to section $M(\hat{s})$ as the search space for the MLE p^*. Since each section $M(\hat{s})$ is affinely isomorphic to the parameter space Ξ, we can introduce a dualistic structure $(\tilde{g}, \tilde{\nabla}^{(e)}, \tilde{\nabla}^{(m)})$ on Ξ in the following way. Firstly, we identify the metric $\tilde{g}$ with the Fisher metric g restricted on $M(\hat{s})$, that is,

$$\begin{aligned}
\tilde{g}_{(\hat{s},\xi)}\left(\frac{\partial}{\partial \xi_\alpha^{(a)}}, \frac{\partial}{\partial \xi_\beta^{(b)}}\right) &= \frac{\partial \eta_{\langle a',\alpha' \rangle}}{\partial \xi_\alpha^{(a)}} \frac{\partial \eta_{\langle b',\beta' \rangle}}{\partial \xi_\beta^{(b)}} g_{(s,\xi)}\left(\frac{\partial}{\partial \eta_{\langle a',\alpha' \rangle}}, \frac{\partial}{\partial \eta_{\langle b',\beta' \rangle}}\right)\Bigg|_{s=\hat{s}} \\
&= s^{(a)} s^{(b)} \frac{\partial^2 \varphi(\eta)}{\partial \eta_{\langle a,\alpha \rangle} \partial \eta_{\langle b,\beta \rangle}}\Bigg|_{s=\hat{s}} \\
&= \delta_{ab}\, \hat{s}^{(a)} \left(\frac{1}{\xi_d^{(a)}} + \frac{\delta_{\alpha\beta}}{\xi_\alpha^{(a)}}\right),
\end{aligned}$$

for $a, b \in \{1, \ldots, d+1\}$ and $\alpha, \beta \in \{1, \ldots, d-1\}$, where $\hat{s}^{(d+1)}$ and $\xi_d^{(a)}$ are formally defined as

$$\hat{s}^{(d+1)} := 1 - \sum_{a=1}^{d} \hat{s}^{(a)}, \qquad \xi_d^{(a)} := 1 - \sum_{\alpha=1}^{d-1} \xi_\alpha^{(a)}.$$

Secondly, the mixture connection $\tilde{\nabla}^{(m)}$ on Ξ is defined through the natural affine isomorphism between $M(\hat{s})$ and Ξ. Finally, the dual connection $\tilde{\nabla}^{(e)}$ is defined by the duality

$$\tilde{g}(\tilde{\nabla}_X^{(e)} Y, Z) := X\tilde{g}(Y, Z) - \tilde{g}(Y, \tilde{\nabla}_X^{(m)} Z).$$

Thus, the MLE ξ^* in the parameter space Ξ is interpreted as the $\tilde{\nabla}^{(m)}$-projection from $\hat{\xi}$ to the physical domain B with respect to the metric $\tilde{g}$.

4. Examples

In this section, we present some examples that demonstrate the implication of Proposition 2 as well as the general diagram given in Figure 2.

4.1. When $\dim \mathcal{H} = 2$

Let us first study the simplest case when $\mathcal{H} = \mathbb{C}^2$. A full set of MUBs is given by

$$\begin{aligned}
\left\{|1^{(1)}\rangle, |2^{(1)}\rangle\right\} &= \left\{\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}\right\}, \\
\left\{|1^{(2)}\rangle, |2^{(2)}\rangle\right\} &= \left\{\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-i\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\i\end{pmatrix}\right\}, \\
\left\{|1^{(3)}\rangle, |2^{(3)}\rangle\right\} &= \left\{\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right\}.
\end{aligned}$$

With these bases, the parameter representation (1) becomes

$$\rho = \frac{1}{2}\begin{pmatrix} 1 + x_3 & x_1 - i x_2 \\ x_1 + i x_2 & 1 - x_3 \end{pmatrix},$$

where $x = (x_1, x_2, x_3)$ is the standard Stokes parameter, which is related to $\xi = (\xi_1^{(1)}, \xi_1^{(2)}, \xi_1^{(3)})$ as $x_a = 2\xi_1^{(a)} - 1$ for $a = 1, 2, 3$.

Figure 3 demonstrates how the $\tilde{\nabla}^{(m)}$-projection is realized. Here, the trajectories of $\tilde{\nabla}^{(m)}$-projections that gives the MLE p^* are plotted only on the x_1x_2-plane. The left and right panels correspond to the cases when $N^{(1)} : N^{(2)} = 1 : 1$ and $N^{(1)} : N^{(2)} = 5 : 1$, respectively. The change of ξ_1-coordinate relative to the change of x_2-coordinate along each trajectory is less noticeable in the right panel than in the left panel. This is because a tomography with $N^{(1)}/N^{(2)} = 5$ provides us with more information about x_1-coordinate, relative to x_2-coordinate, as compared with the case when $N^{(1)}/N^{(2)} = 1$.

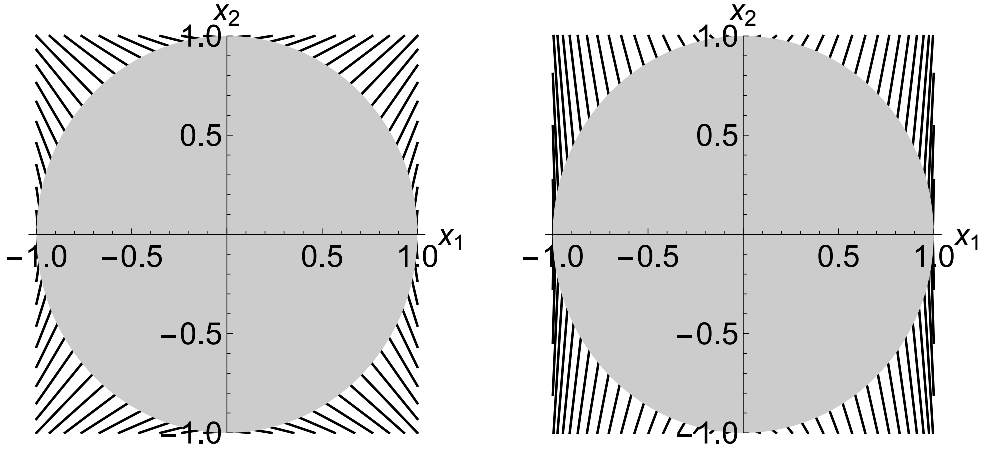

Figure 3. The trajectories of $\tilde{\nabla}^{(m)}$-projections on the Stokes parameter space when $N^{(1)} : N^{(2)} = 1 : 1$ (**left**) and $N^{(1)} : N^{(2)} = 5 : 1$ (**right**). The greyish disk represents the Bloch ball B.

4.2. *When* $\dim \mathcal{H} = 3$

The space $\mathcal{H} = \mathbb{C}^3$ admits a full set of MUBs; for example,

$$\left\{|1^{(1)}\rangle, |2^{(1)}\rangle, |3^{(1)}\rangle\right\} = \left\{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}\right\},$$

$$\left\{|1^{(2)}\rangle, |2^{(2)}\rangle, |3^{(2)}\rangle\right\} = \left\{\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix}1\\\omega\\\omega^2\end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix}1\\\omega^2\\\omega\end{pmatrix}\right\},$$

$$\left\{|1^{(3)}\rangle, |2^{(3)}\rangle, |3^{(3)}\rangle\right\} = \left\{\frac{1}{\sqrt{3}}\begin{pmatrix}\omega\\1\\1\end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix}1\\\omega\\1\end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\\omega\end{pmatrix}\right\},$$

$$\left\{|1^{(4)}\rangle, |2^{(4)}\rangle, |3^{(4)}\rangle\right\} = \left\{\frac{1}{\sqrt{3}}\begin{pmatrix}\omega^2\\1\\1\end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix}1\\\omega^2\\1\end{pmatrix}, \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\\omega^2\end{pmatrix}\right\},$$

where $\omega = (-1 + i\sqrt{3})/2$ is a primitive third root of unity. With these bases, the parameter representation (1) becomes

$$\rho = \begin{pmatrix} \xi_1^{(1)} & a_{12} - ib_{12} & a_{13} - ib_{13} \\ a_{12} + ib_{12} & \xi_2^{(1)} & a_{23} - ib_{23} \\ a_{13} + ib_{13} & a_{23} + ib_{23} & 1 - \xi_1^{(1)} - \xi_2^{(1)} \end{pmatrix},$$

where

$$a_{12} = \frac{1}{2}\left(1 + \xi_1^{(2)} - \xi_1^{(3)} - \xi_2^{(3)} - \xi_1^{(4)} - \xi_2^{(4)}\right),$$
$$a_{13} = \frac{1}{2}\left(-1 + \xi_1^{(2)} + \xi_2^{(3)} + \xi_2^{(4)}\right),$$
$$a_{23} = \frac{1}{2}\left(-1 + \xi_1^{(2)} + \xi_1^{(3)} + \xi_1^{(4)}\right),$$
$$b_{12} = \frac{\sqrt{3}}{6}\left(1 - \xi_1^{(2)} - 2\xi_2^{(2)} + \xi_1^{(3)} - \xi_2^{(3)} - \xi_1^{(4)} + \xi_2^{(4)}\right),$$
$$b_{13} = \frac{\sqrt{3}}{6}\left(-1 + \xi_1^{(2)} + 2\xi_2^{(2)} + 2\xi_1^{(3)} + \xi_2^{(3)} - 2\xi_1^{(4)} - \xi_2^{(4)}\right),$$
$$b_{23} = \frac{\sqrt{3}}{6}\left(1 - \xi_1^{(2)} - 2\xi_2^{(2)} + \xi_1^{(3)} + 2\xi_2^{(3)} - \xi_1^{(4)} - 2\xi_2^{(4)}\right).$$

The physical domain B that corresponds to the state space $\mathcal{S}(\mathbb{C}^3)$ is a compact convex subset of the parameter space $\Xi\,(\subset \mathbb{R}^8)$, and the extreme points of B form an algebraic variety with respect to the parameters

$$\xi = (\xi_1^{(1)}, \xi_2^{(1)}, \xi_1^{(2)}, \xi_2^{(2)}, \xi_1^{(3)}, \xi_2^{(3)}, \xi_1^{(4)}, \xi_2^{(4)}).$$

A numerical example of a $\tilde{\nabla}^{(m)}$-projection that gives the MLE is illustrated in Figure 4, where no probe particle is lost, that is, when

$$\hat{s} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right).$$

In Figure 4, the dot laid outside the greyish region indicates the empirical distribution, i.e., the temporary estimate

$$\hat{\xi} = (0.100, 0.100, 0.066, 0.333, 0.333, 0.333, 0.333, 0.333),$$

and the corresponding MLE is

$$\xi_* = (0.122, 0.122, 0.108, 0.329, 0.299, 0.327, 0.327, 0.299).$$

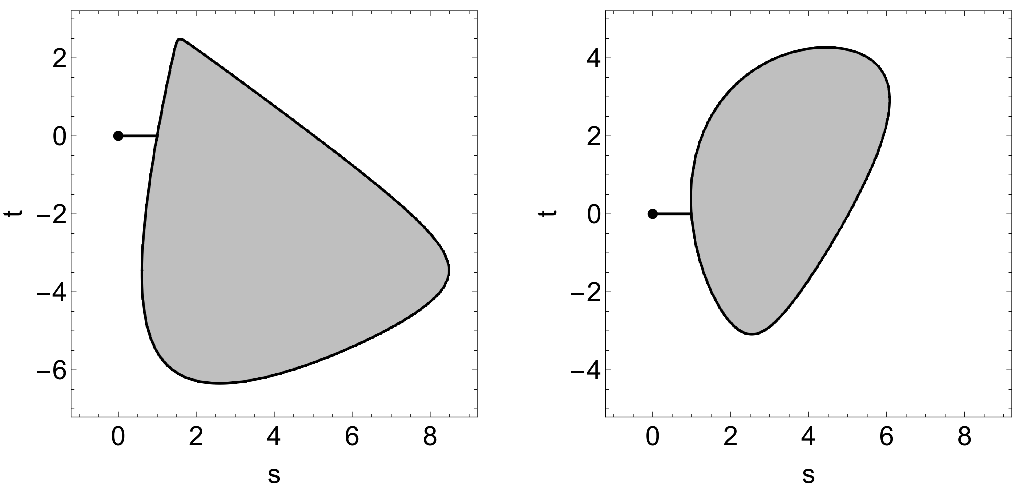

Figure 4. A trajectory of $\tilde{\nabla}^{(m)}$-projection displayed on randomly chosen two-dimensional affine subspaces of Ξ to which both the empirical distribution (marked as a dot) and the MLE belong. The greyish region represents the physical domain B.

Furthermore, the greyish region represents the physical domain B cut by a two-dimensional affine subspace of Ξ specified by the equation

$$\xi = (1-s)\,\hat{\xi} + s\,\xi_* + t\,v.$$

The vector v was chosen randomly under the condition that

$$v \perp \hat{\xi} - \xi_* \quad \text{and} \quad \|v\| = \|\hat{\xi} - \xi_*\|,$$

where the orthogonality $\perp$ and the norm $\|\cdot\|$ are understood relative to the standard Euclidean structure of $\mathbb{R}^8$. In Figure 4, the vector v was taken to be

$$v_1 = (-0.036, -0.038, 0.012, -0.026, -0.038, 0.011, 0.002, 0.005)$$

in the left panel, and

$$v_2 = (0.028, 0.000, -0.006, 0.034, -0.024, -0.022, 0.034, 0.030)$$

in the right panel.

Figure 4 also demonstrates that the sections of the physical domain B show a variety of shapes. Unfortunately, due to this asymmetry of B, we were unable to find a (nontrivial) two-dimensional affine subspace on which every $\tilde{\nabla}^{(m)}$-projection runs. Such a difficulty is in good contrast to the simplest case $\mathcal{H} \simeq \mathbb{C}^2$, where the set B is rotationally symmetric and the $\tilde{\nabla}^{(m)}$-projections can be displayed on any two-dimensional section of B that passes through the origin of B as Figure 3.

4.3. *When* $\dim \mathcal{H} \geq 4$

The space $\mathcal{H} = \mathbb{C}^4$ is also known to admit a full set of MUBs since $\dim \mathcal{H} = 4$ is the second power of the prime number 2; for example [22],

$$\left\{ |1^{(1)}\rangle, |2^{(1)}\rangle, |3^{(1)}\rangle, |4^{(1)}\rangle \right\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\},$$

$$\left\{ |1^{(2)}\rangle, |2^{(2)}\rangle, |3^{(2)}\rangle, |4^{(2)}\rangle \right\} = \left\{ \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} \right\},$$

$$\left\{ |1^{(3)}\rangle, |2^{(3)}\rangle, |3^{(3)}\rangle, |4^{(3)}\rangle \right\} = \left\{ \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ -i \\ -i \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ -1 \\ i \\ i \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ i \\ -i \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ -i \\ i \end{pmatrix} \right\},$$

$$\left\{ |1^{(4)}\rangle, |2^{(4)}\rangle, |3^{(4)}\rangle, |4^{(4)}\rangle \right\} = \left\{ \frac{1}{2}\begin{pmatrix} 1 \\ -i \\ -i \\ -1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ -i \\ i \\ 1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ i \\ i \\ -1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ i \\ -i \\ 1 \end{pmatrix} \right\},$$

$$\left\{ |1^{(5)}\rangle, |2^{(5)}\rangle, |3^{(5)}\rangle, |4^{(5)}\rangle \right\} = \left\{ \frac{1}{2}\begin{pmatrix} 1 \\ -i \\ -1 \\ -i \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ -i \\ 1 \\ i \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ i \\ -1 \\ i \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 \\ i \\ 1 \\ -i \end{pmatrix} \right\}.$$

It is straightforward to calculate the parameter representation (1) of a state $\rho \in \mathcal{S}(C^4)$; however, the corresponding density matrix is rather complicated, and we omit to display it here.

When $\mathcal{H} = \mathbb{C}^6$, or more generally, when $\dim \mathcal{H}$ is not a power of a prime, we do not know whether $\mathcal{H}$ admits a full set of MUBs. Let us touch upon a situation where a Hilbert space $\mathcal{H}$, if it exists, does not admit a full set of MUBs. In this case, there is no measurement basis $M^{(a)}$ that allows a parametrization ξ of the state space $\mathcal{S}(\mathcal{H})$ having a direct connection to the probability distribution of the outcomes as (2). Such a situation could be comparable to the case when the Gell-Mann matrices [23] are used as the measurement basis for estimating an unknown state on $\mathcal{H} = \mathbb{C}^3$. A state $\rho \in \mathcal{S}(\mathbb{C}^3)$ is represented as

$$\rho = \rho_x := \frac{1}{3}\left(I + \sqrt{3}\sum_{i=1}^{8} x_i \lambda_i\right),$$

where $\lambda_1, \ldots, \lambda_8$ are the Gell-Mann matrices, and $x = (x_1, \ldots, x_8)$ is a set of real parameters. The physical domain

$$B = \left\{ x \in \mathbb{R}^8 \,\middle|\, \rho_x \geq 0 \right\}$$

forms a compact convex subset of the unit ball in $\mathbb{R}^8$. With the state ρ_x, the probability distribution of obtaining the eigenvalues $(-1, 0, 1)$ of the observable λ_1 is

$$\left(\frac{1 - \sqrt{3}x_1 + x_8}{3}, \frac{1 - 2x_8}{3}, \frac{1 + \sqrt{3}x_1 + x_8}{3}\right),$$

while the probability distribution of obtaining the eigenvalues $(-1, 0, 1)$ of the observable λ_2 is

$$\left(\frac{1 - \sqrt{3}x_2 + x_8}{3}, \frac{1 - 2x_8}{3}, \frac{1 + \sqrt{3}x_2 + x_8}{3}\right).$$

Note that the probability of obtaining the eigenvalue 0 of λ_1 is identical to that of λ_2. However, in a randomized estimation scheme in which λ_i is measured N_i times, the frequency of obtaining the eigenvalue 0 of λ_1 would be different from that of λ_2. Consequently, one cannot assign a consistent temporary estimate $\hat{x}_8$ for the parameter x_8 in that case. Put differently, the empirical distribution $\hat{q}_N$ on the extended outcome space Ω does not in general have a coordinate representation (4). Thus, the existence of a full set of MUBs is crucial in our analysis.

5. Concluding Remarks

In the present paper, we explored an information geometrical structure of the randomized quantum state tomography, assuming that the Hilbert space under consideration admits a full set of MUBs. We first introduced a classical statistical model $\{p_{(s,\xi)}\}_{s,\xi}$ on an extended sample space Ω, and found that the probability simplex $\mathcal{P}(\Omega)$ was decomposed into mutually orthogonal dualistic foliation (Proposition 1). We then clarified that this geometrical structure had a statistical importance in estimating the coordinate ξ of an unknown quantum state $\rho(\xi)$ under the existence of the nuisance parameter s (Proposition 2). This result gave a generalized insight into the $\nabla^{(m)}$-projection interpretation for the MLE in that a similar interpretation was still valid for the randomized quantum state tomography by changing the standard Fisher metric into a deformed one. It also provided us with a new, convenient way of data processing in the actual quantum state tomography that may involve unexpected probe particle loss.

It should be noted that the existence of a full set of MUBs ensures the parametrization (1) of the quantum state space $\mathcal{S}(\mathcal{H})$. Such a parametrization is distinctive in that it enables a direct correspondence between the parameter space and the probability simplex, realizing the coordinate representation (4) of the empirical distribution $\hat{q}_N$. Thus, the use of a full set of MUBs is crucial in our analysis. Nevertheless, it is often the case that the Hilbert space under consideration takes the form $\mathcal{H} \simeq (\mathbb{C}^p)^{\otimes n}$ for $p = 2$ or 3 because qubits or qutrits are often regarded as building blocks of various quantum protocols. Therefore, the existence of a full set of MUBs would not be too strong a requirement in applications.

Author Contributions: The authors contributed equally to this work.

Funding: The present study was supported by JSPS KAKENHI Grant Numbers JP22340019 and JP17H02861.

Acknowledgments: The authors are grateful to Ryo Okamoto and Shigeki Takeuchi for helpful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MUBs	mutually unbiased bases
PVM	projection-valued measure
MLE	maximum likelihood estimate

Appendix A. Information Geometry of MLE

Let $\mathcal{P}(\Omega)$ denote the set of probability distributions on a finite sample space Ω, i.e.,

$$\mathcal{P}(\Omega) := \left\{ p : \Omega \to \mathbb{R} \;\middle|\; p(\omega) > 0 \text{ for all } \omega \in \Omega, \text{ and } \sum_{\omega \in \Omega} p(\omega) = 1 \right\}.$$

This set may be identified with the $(|\Omega| - 1)$-dimensional (open) simplex, where $|\Omega|$ denotes the number of elements in Ω, and thus it is sometimes referred to as the *probability simplex* on Ω. The set $\mathcal{P}(\Omega)$ is also regarded as a *statistical manifold* endowed with the *dualistic structure* $(g, \nabla^{(e)}, \nabla^{(m)})$, where g is the *Fisher metric*, and $\nabla^{(e)}$ and $\nabla^{(m)}$ are the *exponential* and *mixture connections* [11–13].

Suppose that the state of the physical system at hand belongs to a (closed) subset $\mathcal{M}$ of $\mathcal{P}(\Omega)$, but we do not know which is the true state. We further assume that the probability distributions of $\mathcal{M}$ are faithfully parametrized by a finite dimensional parameter θ as

$$\mathcal{M} = \{p_\theta(\omega) \mid \theta \in \Theta\}.$$

In this case, $\mathcal{M}$ is called a *parametric model*, and our task is to estimate the true value of the parameter θ that specifies the true state. Suppose that, by n independent experiments, we have obtained data $(x_1, x_2, \ldots, x_n) \in \Omega^n$. This information is compressed into the *empirical distribution*, an element of $\mathcal{P}(\Omega)$ defined by

$$\begin{aligned} \hat{q}_n(\omega) &:= \frac{\text{Number of occurrences of } \omega \text{ in data } (x_1, x_2, \ldots, x_n)}{n} \\ &= \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}(\omega) \end{aligned}$$

for each $\omega \in \Omega$, where $\delta_{x_i}(\omega)$ is the Kronecker delta. If $\hat{q}_n$ belongs to the model $\mathcal{M}$, then we have an estimate $\hat{\theta}_n$ that satisfies $p_{\hat{\theta}_n} = \hat{q}_n$. However, the empirical distribution $\hat{q}_n$ does not always belong to the model $\mathcal{M}$. When $\hat{q}_n \notin \mathcal{M}$, we need to find an alternative estimate from the data. A canonical method of finding an alternative estimate $p_{\hat{\theta}_n} \in \mathcal{M}$ is the *maximum likelihood method*, in which one seeks the maximizer of the likelihood function

$$\theta \longmapsto p_\theta(x_1) p_\theta(x_2) \ldots p_\theta(x_n),$$

within the domain Θ of the parameter θ, so that

$$\hat{\theta}_n := \underset{\theta\in\Theta}{\arg\max}\,\{p_\theta(x_1)p_\theta(x_2)\cdots p_\theta(x_n)\}\,.$$

We can rewrite this relation as follows.

$$\begin{aligned}
\hat{\theta}_n &= \underset{\theta\in\Theta}{\arg\max}\ \frac{1}{n}\sum_{i=1}^{n}\log p_\theta(x_i)\\
&= \underset{\theta\in\Theta}{\arg\max}\ \sum_{\omega\in\Omega}\hat{q}_n(\omega)\log p_\theta(\omega)\\
&= \underset{\theta\in\Theta}{\arg\min}\ \sum_{\omega\in\Omega}\hat{q}_n(\omega)\,\{\log\hat{q}_n(\omega)-\log p_\theta(\omega)\}\\
&= \underset{\theta\in\Theta}{\arg\min}\ D(\hat{q}_n\|p_\theta),
\end{aligned}$$

where

$$D(q\|p) := \sum_{\omega\in\Omega} q(\omega)\log\frac{q(\omega)}{p(\omega)}$$

is the *Kullback-Leibler divergence* from q to p. In other words, the maximum likelihood estimate (MLE) $p_{\hat{\theta}_n}$ is the point on $\mathcal{M}$ that is "closest" from the empirical distribution $\hat{q}_n$ as measured by the Kullback-Leibler divergence:

$$p_{\hat{\theta}_n} = \underset{p\in\mathcal{M}}{\arg\min}\ D(\hat{q}_n\|p).$$

Due to the generalized Pythagorean theorem, the MLE is geometrically understood as the $\nabla^{(m)}$-projection from $\hat{q}_n$ to $\mathcal{M}$ or its boundary, as illustrated in Figure A1.

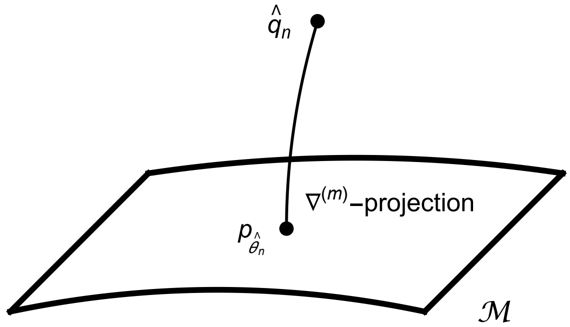

Figure A1. The maximum likelihood estimate $p_{\hat{\theta}_n}$ is the minimizer of the function $p\mapsto D(\hat{q}_n\|p)$ with respect to $p\in\mathcal{M}$, and is also understood as the $\nabla^{(m)}$-projection from the empirical distribution $\hat{q}_n$ to $\mathcal{M}$ or its boundary.

Appendix B. Parametrization of $\mathcal{S}(\mathcal{H})$

Suppose that the Hilbert space $\mathcal{H}\simeq\mathbb{C}^d$ under consideration admits a full set of MUBs

$$\left\{|\alpha^{(a)}\rangle\right\}_{\alpha\in\{1,\dots,d\}},\qquad (a=1,\dots,d+1).$$

For each $a\in\{1,\dots,d+1\}$, let

$$M^{(a)} := \left\{M_1^{(a)},\dots,M_d^{(a)}\right\} = \left\{|1^{(a)}\rangle\langle 1^{(a)}|,\dots,|d^{(a)}\rangle\langle d^{(a)}|\right\}.$$

Then, the operators

$$\left\{M_\alpha^{(a)}-\frac{I}{d}\right\}_{(a,\alpha)\in\{1,\dots,d+1\}\times\{1,\dots,d-1\}}$$

are linearly independent, spanning the space of selfadjoint operators with zero trace. This is easily seen from the orthogonality relation:

$$\mathrm{Tr}\left(M_\alpha^{(a)}-\frac{I}{d}\right)\left(M_\beta^{(b)}-\frac{I}{d}\right)=\delta_{ab}\left(\delta_{\alpha\beta}-\frac{1}{d}\right).$$

Thus, given $\rho\in\mathcal{S}(\mathcal{H})$, the operator $\rho-(I/d)$ is uniquely expanded as

$$\rho-\frac{I}{d}=\sum_{a=1}^{d+1}\sum_{\alpha=1}^{d-1}x_\alpha^{(a)}\left(M_\alpha^{(a)}-\frac{I}{d}\right),$$

where $x_\alpha^{(a)}$ are real numbers. We can regard $x_\alpha^{(a)}$ as a coordinate system of the state space $\mathcal{S}(\mathcal{H})$. When $d=2$, this is identical to the Stokes parametrization, up to a factor of 2.

Now, let us change the coordinate system $x_\alpha^{(a)}$ into $\xi_\alpha^{(a)}$ as

$$x_\alpha^{(a)}=\xi_\alpha^{(a)}+\left(\sum_{\beta=1}^{d-1}\xi_\beta^{(a)}\right)-1.$$

We then arrive at the parametrization (1), i.e.,

$$\rho=\sum_{a=1}^{d+1}\left\{\sum_{\alpha=1}^{d-1}\xi_\alpha^{(a)}M_\alpha^{(a)}+\left(1-\sum_{\beta=1}^{d-1}\xi_\beta^{(a)}\right)M_d^{(a)}\right\}-I.$$

This parametrization is useful in our analysis because it gives a direct connection to the probability distribution of outcomes of the measurement $M^{(a)}$ as

$$p_\alpha^{(a)}:=\mathrm{Tr}\,\rho M_\alpha^{(a)}=\begin{cases}\xi_\alpha^{(a)}, & \text{for } \alpha=1,\ldots,d-1,\\ 1-\sum_{\beta=1}^{d-1}\xi_\beta^{(a)}, & \text{for } \alpha=d.\end{cases}$$

References

1. Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information*; Cambridge University Press: Cambridge, UK, 2000.
2. Lehmann, E.L.; Casella, G. *Theory of Point Estimation*, 2nd ed.; Springer: New York, NY, USA, 1998.
3. Helstrom, C.W. *Quantum Detection and Estimation Theory*; Academic Press: New York, NY, USA, 1976.
4. Holevo, A.S. *Probabilistic and Statistical Aspects of Quantum Theory*; North-Holland: Amsterdam, The Netherlands, 1982.
5. Hradil, Z. Quantum-State Estimation. *Phys. Rev. A* **1997**, *55*, R1561–R1564. [CrossRef]
6. Banaszek, K.; D'Ariano, G.M.; Paris, M.G.A.; Sacchi, M.F. Maximum-likelihood estimation of the density matrix. *Phys. Rev. A* **1999**, *61*, 010304. [CrossRef]
7. Hradil, Z.; Summhammer, J.; Badurek, G.; Rauch, H. Reconstruction of the spin state. *Phys. Rev. A* **2000**, *62*, 014101. [CrossRef]
8. James, D.F.V.; Kwiat, P.G.; Munro, W.J.; White, A.G. Measurement of qubits. *Phys. Rev. A* **2001**, *64*, 052312. [CrossRef]
9. De Burgh, M.D.; Langford, N.K.; Doherty, A.C.; Gilchrist, A. Choice of measurement sets in qubit tomography. *Phys. Rev. A* **2008**, *78*, 052122. [CrossRef]
10. Blune-Kohout, R. Optimal, reliable estimation of quantum states. *New J. Phys.* **2010**, *12*, 043034. [CrossRef]
11. Amari, S.-I.; Nagaoka, H. *Methods of Information Geometry*; Translations of Mathematical Monographs 191; AMS and Oxford: Charles Street, RI, USA, 2000.

12. Amari, S.-I. *Differential-Geometrical Methods in Statistics*; Lecture Notes in Statistics 28; Springer: Berlin, Germany, 1985.
13. Murray, M.K.; Rice, J.W. *Differential Geometry and Statistics*; Chapman & Hall: London, UK, 1993.
14. Fujiwara, A.; Yamagata, K. Data processing for qubit state tomography: An information geometric approach. *arXiv* **2016**, arXiv:1608.07983.
15. Fraïsse, J.M.E.; Braun, D. Quantum channel-estimation with particle loss: GHZ versus W states. *Quantum Meas. Quantum Metrol.* **2016**, *3*, 53. [CrossRef]
16. Durt, T.; Englert, B.-G.; Bengtsson, I.; Życzkowski, K. On mutually unbiased bases. *Int. J. Quantum Inf.* **2010**, *8*, 535–640. [CrossRef]
17. Yuan, H.; Zhou, Z.; Guo, G. Quantum state tomography via mutually unbiased measurements in driven cavity QED systems. *New J. Phys.* **2016**, *18*, 043013. [CrossRef]
18. Wootters, W.K.; Fields, B.D. Optimal state-determination by mutually unbiased measurements. *Ann. Phys.* **1989**, *191*, 363–381. [CrossRef]
19. Bengtsson, I. Three ways to look at mutually unbiased bases. *arXiv* **2006**, arXiv:quant-ph/0610216.
20. Ivonovic, I.D. Geometrical description of quantal state determination. *J. Phys. A* **1981**, *14*, 3241–3245. [CrossRef]
21. Yamagata, K. Efficiency of quantum state tomography for qubits. *Int. J. Quantum Inform.* **2011**, *9*, 1167–1183. [CrossRef]
22. Klappenecker, A.; Rötteler, M. Constructions of mutually unbiased bases. *arXiv* **2003**, arXiv:quant-ph/0309120.
23. Gell-Mann, M. Symmetries of baryons and mesons. *Phys. Rev.* **1962**, *125*, 1067. [CrossRef]