



Article An Information Quantity in Pure State Models

Fuyuhiko Tanaka ^{1,2}

- ¹ Graduate School of Engineering Science, Osaka University, Osaka 560-8531, Japan; ftanaka@sigmath.es.osaka-u.ac.jp
- ² Center for Quantum Information and Quantum Biology, Osaka University, Osaka 560-8531, Japan

Abstract: When we consider an error model in a quantum computing system, we assume a parametric model where a prepared qubit belongs. Keeping this in mind, we focus on the evaluation of the amount of information we obtain when we know the system belongs to the model within the parameter range. Excluding classical fluctuations, uncertainty still remains in the system. We propose an information quantity called purely quantum information to evaluate this and give it an operational meaning. For the qubit case, it is relevant to the facility location problem on the unit sphere, which is well known in operations research. For general cases, we extend this to the facility location problem in complex projective spaces. Purely quantum information reflects the uncertainty of a quantum system and is related to the minimum entropy rather than the von Neumann entropy.

Keywords: minimum entropy; facility location problem; quantum information; complex projective space

MSC: 81P17; 81P05; 81P45; 90B99; 53A20

1. Introduction

Building a large-scale quantum computer remains challenging, and there are many problems to be solved. For example, the performance of error correction depends very strongly on how coherent the noise process is [1], and experimenters need to improve the quantum computing system through analysis of the physical noise [2]. When we prepare an imperfect quantum computing system, it is important to specify the noise based on a suitable error model. The process of understanding the physical error model for a prepared system corresponds to that of obtaining a certain amount of information on the system. In the present paper, keeping this in mind, we consider how to evaluate such information *without any entropic concept*.

An overly simple example is given by a purely rotation error model [2] with a parameter range. Let the ideal qubit state be $|0\rangle$ and the error model be $e^{-i\theta Y}|0\rangle$, where $Y = -i(|0\rangle\langle 1| - |1\rangle\langle 0|)$ and θ denotes the parameter to be specified. Then the parameter range reflects the information we have on the system. We have more information when we know $-\epsilon \leq \theta \leq \epsilon$ than when we know $-2\epsilon \leq \theta \leq 2\epsilon$.

However, what if we consider a more complicated situation such as $|\theta_1, \theta_2\rangle = e^{-i\theta_2 Z}$ $e^{-i\theta_1 Y}|0\rangle$, where $Z = |0\rangle\langle 0| - |1\rangle\langle 1|$, with the parameter range of, say, $0 \le \theta_1 \le \pi/4$ and $0 \le \theta_2 \le \pi/4$? How do we compare the range with the parameter range $0 \le \theta_1 \le 2\pi/5$ and $\theta_2 = 0$? To simplify the problem, let us adopt discrete models. We consider two situations: First, the qubit is described by one of the candidate pure states, $|0\rangle$, $|+\rangle$, $|Y+\rangle$, where $|+\rangle = |\theta_1 = \pi/4$, $\theta_2 = 0\rangle$ and $|Y+\rangle = |\theta_1 = \pi/4$, $\theta_2 = \pi/4\rangle$. Second, it is described by one of the candidate pure states $|0\rangle$, $\frac{|0\rangle+\sqrt{2}|1\rangle}{\sqrt{3}}$, $|+\rangle$. Then, which state's information is greater? Or equivalently, which uncertainty is larger? We will give a definite answer in the present article (See Section 8).

Let us describe our problem in a slightly more formal way before going into detail. Let the quantum system be described in a separable Hilbert space and a subset of pure states $\mathcal{M} = \{\psi_1, \psi_2, ...\}$ be given. We call the subset \mathcal{M} a (*pure-state*) model. Suppose we



Citation: Tanaka, F. An Information Quantity in Pure State Models. *Entropy* 2022, 24, 541. https:// doi.org/10.3390/e24040541

Academic Editors: Durdu Guney and David Petrosyan

Received: 28 March 2022 Accepted: 10 April 2022 Published: 12 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). know that the quantum system is described by one of the pure states in the model. Then we evaluate the amount of information obtained by knowing the model.

We write the model as a countable set for simplicity, but a model might consist of an uncountable set of pure states, e.g., the above parametric pure state $|\theta_1, \theta_2\rangle$. Another continuous model would be a wavefunction with a certain continuous parameter { $\varphi(x-s) : -\infty < s < \infty$ } (see, e.g., Holevo [3]).

Our problem is closely related to so-called quantum estimation, but the above type of problem has still not been investigated. In quantum state estimation [4–9] or quantum state discrimination [6,10–13], for a given model, we find an optimal quantum measurement to extract information on the quantum state and choose the true state in the model from observation. This has been a typical problem and has been investigated by many authors. In quantum information experiments, quantum tomography has been also discussed [14–17]. As far as the author knows, these studies do not refer to the comparison of several models in terms of information quantity.

In our setting, we focus on the information that we obtain *before* preparation for a measurement. As we see later, we clearly obtain a certain amount of information other than the dimension of the Hilbert space.

We consider only pure state models so that we can neglect classical fluctuation. As we shall see later, there is no classical counterpart for such information. In other words, we calibrate so that such classical information becomes zero. If a positive amount of information remains under the calibration, then we expect that it reflects the truly quantum information. We do not have a proper name for this information, and we call it *model information* or *information of the model* tentatively. It would become an alternative to the usual entropy.

In the next section, we provide a rough idea of how to define model information and present pure state models as examples. Then, we will formulate a pure state model and define the representative quantum state for it in a rigorous manner. In Section 4, we describe the equivalence between the problem of finding the representative quantum state and finding the minimax facility location on the sphere in operations research. In Section 5, we introduce the purely quantum information of the model and calculate it in several examples. We also describe the relationship of entropic concepts to our result and extension to infinite-dimensional Hilbert space in Section 7. Finally, concluding remarks are given in Section 8.

2. Rough Idea on Defining Model Information

2.1. Preliminary Considerations

In this section, we describe a rough idea of how we evaluate model information. First, we recall classical information theory. Suppose that Alice picks a three-letter word $w_1w_2w_3$, where $w_1, w_2, w_3 \in \{a, b, c, ..., z\}$ and we set $\mathcal{M} = \{a, b, c, ..., y\}$. If Bob knows $w_1 \in \mathcal{M}$, Bob does not feel that he obtains much information. However, if $w_1 \in \mathcal{M}' = \{a, e, i, o, u\}$, then Bob feels that he obtains more information on the word Alice picks.

The above situation corresponds to a commutative case in quantum theory. Keeping this in mind, let us consider the model information in the quantum system. We assume that Bob already knows that the quantum system is described in a *d*-dimensional Hilbert space. Since information quantity is a relative concept, let us compare two models. Let the first model consist of a *d*-dimensional orthonormal basis, i.e., $\mathcal{M} = \{e_1, \ldots, e_d\}$ and the second model consist of $\{e_1, \ldots, e_k\}$ (k << d). At least we can say that the second model gives more significant information to Bob than the first model. This is because the quantum state is in a proper subspace.

Now we tackle the case where some quantum states are nonorthogonal. For simplicity, we set d = 2 and consider the following models:

$$\begin{split} \mathcal{M}_{1} &= \{|+\rangle, |-\rangle\},\\ \mathcal{M}_{2} &= \{|0\rangle, |+\rangle\},\\ \mathcal{M}_{3} &= \{|0\rangle, \frac{|0\rangle + \sqrt{2}|1\rangle}{\sqrt{3}}, |+\rangle\},\\ \mathcal{M}_{4} &= \{|0\rangle, |+\rangle, |Y+\rangle\}, \end{split}$$

where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$, $|Y+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle)$. In explicit calculations, we set $|0\rangle = (1,0)^{\top}, |1\rangle = (0,1)^{\top}$.

Suppose that we know that the quantum state φ is one of the candidate states in \mathcal{M}_2 (hereinafter, we write $\varphi \in \mathcal{M}_2$ for simplicity). Perhaps we agree that the information is more than $\varphi \in \mathcal{M}_3$ and $\varphi \in \mathcal{M}_4$. Then, which is more informative, $\varphi \in \mathcal{M}_3$ or $\varphi \in \mathcal{M}_4$? Both models consist of three nonorthogonal state vectors. Likewise, which is more informative, $\varphi \in \mathcal{M}_1$ or $\varphi \in \mathcal{M}_2$? In the present article, we consider how to quantitatively evaluate the information obtained when Bob knows that the quantum state φ belongs to a model \mathcal{M} .

2.2. Full Rank Condition

In order to avoid technical difficulties, we give one important assumption here. Let us define the *rank* of a model as

$$\operatorname{cank}\mathcal{M} = \operatorname{dim}\operatorname{span}_{\mathbf{C}}\{\psi_1, \psi_2, \dots\}.$$

We assume that the rank of a model is equal to the dimension of the Hilbert space, i.e., $d = \operatorname{rank} \mathcal{M}$. We call it the *full rank condition*. The full rank condition implies that there exists no complementary subspace that is orthogonal to every state vector in the model \mathcal{M} .

2.3. Rough Idea on Defining Model Information

Under the full rank condition, we consider the case where a model has considerable information on the quantum system. Suppose that we are given the following model:

$$\mathcal{M} = \{ |e_1\rangle, \sqrt{1-\epsilon} |e_1\rangle + \sqrt{\epsilon} |e_2\rangle, \dots, \sqrt{1-\epsilon} |e_1\rangle + \sqrt{\epsilon} |e_d\rangle \}$$

 $(0 < \epsilon << 1)$. While this satisfies the full rank condition, clearly all candidate quantum states are approximately in the same direction as $|e_1\rangle$.

Then, the quantum system φ is approximately described by a representative state vector $|e_1\rangle$. When $\epsilon \to 0$ but $\epsilon \neq 0$, the model information is expected to increase.

From the above discussion, we find that the information quantity associated with $\varphi \in \mathcal{M}$ is completely different from the number of elements, $|\mathcal{M}|$. Rather, a certain scale or a size of the model \mathcal{M} should be included in the definition of the model information.

Along the lines of the above rough idea, we discuss in the next section:

- (a) How to determine a representative state vector for a given model \mathcal{M} ;
- (b) One definition of the model information;
- (c) The relationship with the concept of entropy.

We emphasize that all of these have no classical counterpart and thus it might be difficult to understand them. Before going into detail, we shall give an overview of each item here.

For (a), we consider maximin overlap between quantum states and define the representative quantum state of a model. Mathematically speaking, it is regarded as a variant of the facility location problem on the unit sphere [18,19], which appears in operations research. In operations research, many authors have developed algorithms on the facility location problem. In particular, finding the minimax solution is our concern. For a finite model ($|\mathcal{M}| < \infty$), we present a naive algorithm to find the representative quantum state for a model using this consideration. In order to consider item (b), we introduce an imaginary two-person game called the quantum detection game. Bob benefits from the information of a given model to obtain a higher score than Alice. The value of the game, which is determined by a least favorable prior [20] in this game, defines one information quantity related to the model \mathcal{M} .

In (c), we compare our method with the formal analogue based on the von Neumann entropy. Later we will see that the newly proposed information quantity is related to the minimum entropy [21,22] rather than the von Neumann entropy.

3. Basic Definitions

3.1. Definition of Pure State Models and Assumptions

In the present paper, let \mathcal{H} be a *d*-dimensional Hilbert space. (*d* could be ∞). We call a finite-dimensional parametric family of quantum pure states

$$\mathcal{M} = \{ \psi_{\theta} \in \mathcal{H} : \| \psi_{\theta} \| = 1, \theta \in \Theta \subset \mathbf{R}^m \}$$

a *quantum statistical model of pure states* or briefly a (*pure state*) *model*. Note that $||f||^2 = \langle f|f \rangle$. Basically, the parameter set Θ is a compact subset of finite-dimensional Euclidean space. We assume the following two conditions:

- (1) Identifiability: $\theta \neq \theta' \Rightarrow |\psi_{\theta}\rangle \langle \psi_{\theta}| \neq |\psi_{\theta'}\rangle \langle \psi_{\theta'}|$. Conventionally, we only consider quantum states up to the (global) phase factor below, and we often identify a pure state $|\psi\rangle$ with a density operator $|\psi\rangle \langle \psi|$.
- (2) Continuity: For every sequence $\{\theta_n\}_{n=1}^{\infty} \subset \Theta$ and $\theta \in \Theta$,

$$\theta_n o heta \Rightarrow \||\psi_{\theta_n}\rangle\langle\psi_{\theta_n}| - |\psi_{ heta}\rangle\langle\psi_{ heta}|\|_{\infty} o 0$$

holds. ($||A||_{\infty}$ denotes the operator norm, i.e., $||A||_{\infty} = \sup_{\|\varphi\|=1} ||A\varphi||$).

For simplicity, we often consider a finite set of parameters, $\Theta = \{\theta_1, \ldots, \theta_k\}$. Then, ψ_{θ_j} is denoted by ψ_i . We often call it a *discrete model*, which is written as $\mathcal{M} = \{\psi_1, \ldots, \psi_k\}$.

3.2. Preliminary Results

In the present paper, we introduce the information of a model \mathcal{M} . Although the formal definition is given in Section 5, we need several concepts to understand them analytically and geometrically.

In this section, we introduce the most fundamental concept, the representative quantum state of a model. We shall give a rough idea for when dim $\mathcal{H} = 2$ and $|\mathcal{M}| = 2$. Specifically, we set $\mathcal{M} = \{\psi_1, \psi_2\}$ with $||\psi_1|| = ||\psi_2|| = 1$. When two quantum states are close to each other, $\psi_1 \simeq \psi_2$, it is natural to consider that a representative quantum state of model \mathcal{M} should be a "midpoint between two quantum states". We often identify the state vector with the point on the whole pure states specified by the vector.

Mathematically, we may try to define the point as the point φ equidistant between ψ_1 and ψ_2 such that

$$\langle \varphi | \psi_1 \rangle | = |\langle \varphi | \psi_2 \rangle| \tag{1}$$

holds.

However, the above equidistance condition does not determine the point φ generally. Thus, we maximize the above "overlap" under the condition (1). Then we obtain an explicit formula for the representative point of a model,

$$|\varphi_{rep}\rangle = \frac{|\psi_1\rangle + e^{-i\delta}|\psi_2\rangle}{\sqrt{2(1+|\gamma|)}}, \qquad \gamma = \langle\psi_1|\psi_2\rangle, \delta = \arg\gamma$$
(2)

where δ satisfies $\gamma = |\gamma|e^{i\delta}$ and then the maximum overlap is given by $\max_{\varphi} |\langle \varphi | \psi_1 \rangle| = \frac{1+|\gamma|}{2}$.

Next, we consider the case where $|\mathcal{M}| = 3$. Let us take \mathcal{M}_3 and \mathcal{M}_4 introduced in the previous section:

In \mathcal{M}_4 , the above idea applies, i.e., we find the quantum state to maximize the overlap

$$|\langle \varphi | \psi_1 \rangle| = |\langle \varphi | \psi_2 \rangle| = |\langle \varphi | \psi_3 \rangle|. \tag{3}$$

Up to the global phase, we set φ as

$$|\varphi\rangle = {\cos \alpha \choose \sin \alpha \, e^{i\delta}}, \ 0 \le \alpha \le \pi/2, \ 0 \le \delta < 2\pi.$$
(4)

Then, we obtain an explicit solution satisfying (3), $\cos \alpha = \sqrt{\frac{3+\sqrt{3}}{6}}$, $\delta = \frac{\pi}{4}$ after lengthy but straightforward algebra. (See also Section 4.1.3).

However, we find no solution satisfying Equation (3) in \mathcal{M}_3 . We need a more careful treatment. First, we fix an arbitrary quantum state φ and consider the set of numbers r satisfying $|\langle \varphi | \psi_j \rangle| \ge r$, j = 1, 2, 3. The condition assures that the overlap between φ and an arbitrary quantum state in \mathcal{M} is not less than r. For each φ , the maximum of r is equal to $\min\{|\langle \varphi | \psi_j \rangle| : j = 1, 2, 3\}$.

We consider that the larger the overlap gets, the more suitable φ becomes as a representative quantum state of the model M. Thus, we maximize *r* as a function of φ .

It is convenient for explicit calculation to use the squared overlap (i.e., Fidelity), $\max_{\varphi} \min_{\psi \in \mathcal{M}_3} |\langle \varphi | \psi \rangle|^2$, and we regard $\varphi_{rep} = \arg \max r(\varphi; \mathcal{M}_3)$ as a representative quantum state of the model \mathcal{M}_3 . Based on the above idea, we will give a more formal definition of the representative quantum state in the next subsection.

3.3. Representative Quantum State

Now we are ready to define the representative quantum state of a given model \mathcal{M} formally. We adopt the distance $d_F(\varphi, \varphi') = 1 - F(\varphi, \varphi') = 1 - |\langle \varphi | \varphi' \rangle|^2$ rather than the overlap.

Definition 1. Let a model \mathcal{M} be given. When a quantum state φ_{rep} satisfies

$$\inf_{\varphi} \sup_{\psi \in \mathcal{M}} d_F(\varphi, \psi) = \sup_{\psi \in \mathcal{M}} d_F(\varphi_{rep}, \psi),$$
(5)

 φ_{rep} is called a representative quantum state of the model \mathcal{M} with respect to the distance d_F .

When we emphasize the model \mathcal{M} , we write $\varphi_{rep}(\mathcal{M})$. While the terms max and min are enough for discrete models, using the terms sup and inf is generally inevitable. (see Section 7). We also use a condition equivalent to (5),

$$\sup_{\psi \in \mathcal{M}} d_F(\varphi, \psi) \geq \sup_{\psi \in \mathcal{M}} d_F(\varphi_{rep}, \psi), \qquad \forall \varphi$$

In the above definition, φ_{rep} is also interpreted as the minimax estimate in quantum estimation with *no observation*. Suppose that a parametric family of pure states or countable set of pure states is given. Then we give an estimate, say φ , as the true quantum state without any observation. The error is evaluated by the Fidelity-based quantity, $d_F(\varphi, \psi_{true}) = 1 - |\langle \varphi | \psi_{true} \rangle|^2$. The above representative quantum state is a minimax estimate in this setting.

In the context of quantum estimation, this may seem quite strange because we do not perform any measurement. However, it is not unnatural to consider estimation with no observation. For example, in classical information theory, we infer the outcome of an information source with no observation. For a given parametric model of source code distribution $\{p_{\theta}(x) : \theta \in \Theta\}$, this kind of estimation corresponds to constructing a minimax code [23].

Apart from actual application, quantum estimation with no observation also makes sense theoretically. In a quantum computer, a quantum bit will be processed under a certain quantum gate with an unknown parameter, say, θ , during the computing process. When θ is uncontrollable with a range $[-2\epsilon, \epsilon]$, it might be necessary to estimate the quantum bit. Since there is no reason to estimate $\theta \simeq 0.5\epsilon$, we need a certain formulation to estimate the quantum bit.

We should also mention why we adopt d_F as the distance among several candidates as the closeness measure in our definition. There are two reasons. One is the operational meaning of the quantum detection game, which is explained in Section 5. The other is due to the following property:

Lemma 1. Let a model $\mathcal{M} = \{\psi_{\theta}\}$ be given. Let f be a continuous nondecreasing function on [0,1]. When we adopt the distance $d(\varphi, \psi) = f \circ d_F(\varphi, \psi)$, then the representative quantum state remains the same.

Proof. It is enough to show that for every φ ,

$$f\left\{\sup_{\theta} d_F(\varphi, \psi_{\theta})\right\} = \sup_{\theta} f \circ d_F(\varphi, \psi_{\theta})$$
(6)

holds.

If Equation (6) holds true, then we show the statement in the following way. For every φ , from the definition of φ_{rep} , $\sup_{\theta} d_F(\varphi, \psi_{\theta}) \ge \sup_{\theta} d_F(\varphi_{rep}, \psi_{\theta})$ holds. Since *f* is an increasing function, applying *f* to both sides and using Equation (6) yields $\sup_{\theta} f \circ d_F(\varphi, \psi_{\theta}) \ge \sup_{\theta} f \circ d_F(\varphi_{rep}, \psi_{\theta})$, which implies that φ_{rep} is a representative quantum state with respect to the distance $f \circ d_F$.

Now let us show Equation (6). Let φ be fixed and set $\alpha = \sup_{\theta} d_F(\varphi, \psi_{\theta})$. For every $\epsilon > 0$, due to the continuity of f, there exists $\delta > 0$ such that $|\alpha - x| \le \delta \Rightarrow f(\alpha) \le f(x) + \epsilon$. We take θ_* such that $\alpha \le d_F(\varphi, \psi_{\theta_*}) + \delta$. Then

$$f(\alpha) \leq f(d_F(\varphi, \psi_{\theta_*})) + \epsilon$$

$$\leq \sup_{\varphi} f \circ d_F(\varphi, \psi_{\theta}) + \epsilon$$

Since ϵ is arbitrary, we obtain $f(\alpha) \leq \sup_{\theta} f \circ d_F(\varphi, \psi_{\theta})$.

Next, observe that $f(\alpha) \ge f \circ d_F(\varphi, \psi_{\theta})$ for every θ since $\alpha \ge d_F(\varphi, \psi_{\theta})$. Taking the supremum of RHS with respect to θ , we obtain the converse inequality. Thus, Equation (6) is shown, and the proof is complete. \Box

In Section 5, we shall define the information quantity obtained when we know $\varphi \in \mathcal{M}$, which is denoted by $J(\mathcal{M})$. When we find φ_{rep} , it is shown to be easy to calculate $J(\mathcal{M})$.

Now let us consider the representative quantum state of a two-state model geometrically. Recall that each pure state in a two-dimensional Hilbert space is written in the form (4). If we switch to the Bloch representation, we obtain one-to-one correspondence between $(\alpha, \delta) \leftrightarrow (x, y, z) = (\sin 2\alpha \cos \delta, \sin 2\alpha \sin \delta, \cos 2\alpha)$ on the unit sphere (Bloch sphere). When one pure state is set to $|0\rangle$ (*P*), the distance between the pure state and another pure state specified with (α, δ) (*Q*) is 2α along the shortest path on the Bloch sphere. The shortest path connecting two points *P* and *Q* on the Bloch sphere is the arc along the large circle on the Bloch sphere. The arc is called a *geodesic* connecting *P* and *Q* and the equidistant point *M* on the geodesic from both points is called the *geodesic* midpoint *between P and Q*. The representative quantum state corresponds to the geodesic midpoint. The concept of geodesics on the Bloch sphere is often useful and has been investigated in several works [24–27].

For every pair of independent quantum states ψ_1 and ψ_2 , let us consider a twodimensional subspace span_C{ ψ_1, ψ_2 }. Then each state in the subspace is regarded as a point on the Bloch sphere. By using the Formula (2), we summarize the above statements.

Lemma 2. Let a model \mathcal{M} be given. Then, for every pair of quantum states ψ_1 and ψ_2 with $\|\psi_1\| = \|\psi_2\| = 1$, the geodesic midpoint φ_M is given by

$$|\varphi_M\rangle = \frac{|\psi_1\rangle + e^{-i\delta}|\psi_2\rangle}{\sqrt{2(1+|\gamma|)}}, \qquad \gamma = \langle \psi_1|\psi_2\rangle, \ \delta = \arg\gamma$$
(7)

where δ satisfies $\gamma = |\gamma|e^{i\delta}$ and then arc length α between φ_M and ψ_1 is given by $\alpha = \arccos|\langle \psi_1 | \psi_2 \rangle|$ and arc length between ψ_1 and ψ_2 is 2α .

Understanding the geometry of the unit sphere is very helpful to find the representative quantum state, which is discussed in Section 4.

3.4. Example of a Representative Quantum State: M_3

As a slightly nontrivial example, let us focus on \mathcal{M}_3 and find a representative quantum state. First, we focus on the submodel $\mathcal{N}_3 = \{|0\rangle, \frac{|0\rangle + \sqrt{2}|1\rangle}{\sqrt{3}}\}$. Then its representative quantum state, $\varphi_{rep}(\mathcal{N}_3)$ is the geodesic midpoint of $|0\rangle$ and $\frac{|0\rangle + \sqrt{2}|1\rangle}{\sqrt{3}}$. Using the Formula (7),

$$|\varphi_{rep}(\mathcal{N}_3)\rangle = \frac{(\sqrt{3}+1)|0\rangle + \sqrt{2}|1\rangle}{2\sqrt{(\sqrt{3}+1)}}$$
(8)

is obtained.

Next, we use the following lemma.

Lemma 3. Let a model \mathcal{M} and its submodel $\mathcal{N} \subset \mathcal{M}$ be given. If the representative quantum state φ_{rep} of \mathcal{N} satisfies

$$\sup_{\psi \in \mathcal{N}} d_F(\varphi_{rep}, \psi) = \sup_{\psi \in \mathcal{M}} d_F(\varphi_{rep}, \psi),$$

then φ_{rep} is also the representative quantum state of \mathcal{M} .

Proof. Let φ be an arbitrary quantum state. Since $\mathcal{N} \subset \mathcal{M}$,

$$\sup_{\psi \in \mathcal{M}} d_F(\varphi, \psi) \ge \sup_{\psi \in \mathcal{N}} d_F(\varphi, \psi) \ \ge \sup_{\psi \in \mathcal{N}} d_F(\varphi_{rep}, \psi) \ = \sup_{\psi \in \mathcal{M}} d_F(\varphi_{rep}, \psi),$$

which implies that φ_{rep} is also the representative quantum state of \mathcal{M} . \Box

It is easily seen that $d_F(\varphi_{rep}(\mathcal{N}_3), |+\rangle) < \max_{\psi \in \mathcal{N}_3} d_F(\varphi_{rep}(\mathcal{N}_3), \psi)$. Due to the above Lemma 3, $\varphi_{rep}(\mathcal{N}_3)$ (8) is also the representative quantum state of \mathcal{M}_3 .

4. Facility Location Problem

Mathematically, finding the representative quantum state of a given model is equivalent to finding the minimax facility location for a given demand point in operations research [18].

4.1. Facility Location Problem on the Sphere

Decades ago, Drezner and Wesolowsky [18] considered the facility location problem on the sphere. We briefly summarize their formulation. Suppose that there are *m* demand points with (positive) weights on the unit sphere and our objective is to locate a single facility on the same sphere so as to minimize the weighted sum of distances from the facility to the demand points. Let (ξ_i, ϕ_i) and (ξ, ϕ) denote the locations of the *i*-th demand point and the facility, respectively, in spherical coordinate ($0 \le \xi \le \pi$, $0 \le \phi < 2\pi$). Weights for demand points are denoted as π_i . Without loss of generality, we may take $\sum_i \pi_i = 1$. We obtain the following minimization problem.

$$\min_{\boldsymbol{\xi},\boldsymbol{\phi}}\sum_{i=1}^m \pi_i d_i(\boldsymbol{\xi},\boldsymbol{\phi}),$$

where $d_i(\xi, \phi)$ is the distance between the facility (ξ, ϕ) and the *i*-th demand point (ξ_i, ϕ_i) .

Drezner and Wesolowsky measured distances through the sphere for squared Euclidean distances and they also used the shortest length of arc. Let us denote the shortest length of arc between two points on a sphere with a unit radius by α . Then the squared Euclidean distance is given by $4\sin^2(\alpha/2)$. Both distances are computed by the equation:

$$\cos \alpha = \cos \xi_i \cos \xi + \sin \xi_i \sin \xi \cos(\phi_i - \phi).$$

The interpretation of the problem is as follows. The distance $d_i(\xi, \phi)$ is the transportation cost from a facility (ξ, ϕ) to the *i*-th demand point (ξ_i, ϕ_i) . We regard the relative frequency of each transport as the weight. When we already know the relative frequency, we minimize the objective function $\sum_{i=1}^{m} \pi_i d_i(\xi, \phi)$ with respect to (ξ, ϕ) .

Focusing on the correspondence between a point (ξ, ϕ) on the unit sphere and a complex unit vector

$$|\psi(\xi,\phi)\rangle = \begin{pmatrix} \cos\xi/2\\ e^{i\phi}\sin\xi/2 \end{pmatrix},\tag{9}$$

the problem is completely solved when we adopt the squared Euclidean distance. Let us denote $|\varphi\rangle = |\psi(\xi, \phi)\rangle$ as the location of the facility instead of (ξ, ϕ) . Straightforward calculation yields $d_i(\xi, \phi) = 4\{1 - |\langle \varphi | \psi(\xi_i, \phi_i) \rangle|^2\} = 4d_F(\varphi, \psi(\xi_i, \phi_i))$. Thus, the objective function to be minimized is written as

$$\begin{split} \sum_{i=1}^{m} \pi_{i} d_{i}(\xi, \phi) &= 4 \sum_{i=1}^{m} \pi_{i} \Big\{ 1 - |\langle \varphi | \psi(\xi_{i}, \phi_{i}) \rangle|^{2} \Big\} \\ &= 4 \langle \varphi | \bigg\{ I - \left(\sum_{i=1}^{m} \pi_{i} | \psi(\xi_{i}, \phi_{i}) \rangle \langle \psi(\xi_{i}, \phi_{i}) | \right) \bigg\} | \varphi \rangle \\ &= 4 (1 - \langle \varphi | \rho_{\pi} | \varphi \rangle), \end{split}$$

where in the last line, we set $\rho_{\pi} = \sum_{i=1}^{m} \pi_i |\psi(\xi_i, \phi_i)\rangle \langle \psi(\xi_i, \phi_i)|$. Note that ρ_{π} is positive semidefinite and of trace one, and it is regarded as the *Bayes mixture* (For the definition, see Section 5.1). Then, the minimization problem $\min_{\varphi} \sum_{i=1}^{m} \pi_i d_i(\xi, \phi)$ reduces to finding the maximum of $\langle \varphi | \rho_{\pi} | \varphi \rangle$. This is given by the first eigenvector φ of ρ_{π} . This result [28] agrees with that derived by Drezner and Wesolowsky, who obtained the same result by differentiation with respect to variables (ξ, ϕ) in the context of operations research.

However, what if we have no information on the relative frequency $\{\pi_i\}$ for each demand point? One idea is to take the minimax point. Through the minimax theorem [20], we obtain the associated weight $\pi^* = \{\pi_i^*\}$ such that

$$\min_{\xi,\phi} \sum_{i=1}^{m} \pi_i^* d_i(\xi,\phi) = \min_{\xi,\phi} \max_{i=1,\dots,m} d_i(\xi,\phi)$$
(10)

holds, where $d_i(\xi, \phi)$ is the squared Euclidean distance. (Strictly speaking, the above holds under the condition that the convex hull of all demand points does not include the origin). This π^* is called the *least favorable weight*.

Now we go back to our problem. Let the Hilbert space be two-dimensional, where each quantum state is specified by (ξ, ϕ) in (9). Let a discrete model $\mathcal{M} = \{|\psi(\xi_1, \phi_1)\rangle, \dots, |\psi(\xi_m, \phi_m)\rangle\}$ be given. Each state corresponds to a demand point on the Bloch sphere. The distance d_F corresponds to the transportation cost measured by a constant multiplied by the squared Euclidean distance. Then, finding the representative quantum state φ_{rep} of the model \mathcal{M} is equivalent to finding the minimax facility location specified by φ on the Bloch sphere with the distance d_F .

According to this correspondence, we also know the following fact. When we know the least favorable weight π_* in the problem, we obtain the representative quantum state as the first eigenvector of the Bayes mixture ρ_{π} with respect to π_* [20].

Following the interpretation of the facility location problem on the sphere, we find the representative quantum state of each model in Section 2.1. Since $\sin^2(\alpha/2)$ is a strictly increasing function of α , both minimax points for the two distance measures agree due to Lemma 1.

4.1.1. Example 1: \mathcal{M}_2

Pure states $|+\rangle$ and $|0\rangle$ correspond to the demand points specified by $(\xi, \phi) = (\pi/2, 0)$ and $(\xi, \phi) = (0, 0)$ (North pole) on the Bloch sphere, respectively. Then, the minimax location on the Bloch sphere is specified by $(\xi_M, \phi_M) = (\pi/4, 0)$. Thus, φ_{rep} of \mathcal{M}_2 is given by $|\varphi_{rep}\rangle = (\cos \pi/8, \sin \pi/8)^{\top}$. Figure 1 shows the demand points and the minimax location of the facility on the Bloch sphere.



Figure 1. Configuration of demand points in M_2 and the minimax facility location on the Bloch sphere: The red solid points denote demand points and the green solid point denotes the minimax facility location (the representative quantum state), which is the geodesic midpoint.

4.1.2. Example 2: \mathcal{M}_3

First, the quantum state $\frac{|0\rangle+\sqrt{2}|1\rangle}{\sqrt{3}}$ corresponds to the demand point specified by $(\xi_*, 0)$, where $\cos \xi_*/2 = 1/\sqrt{3}$, $\sin \xi_*/2 = \sqrt{2/3}$. We follow Algorithm Step (1) described in the next subsection. Then, the most distant pair is $(\xi, \phi) = (0, 0)$ (North pole) and $(\xi_*, 0)$. The other point specified by $(\xi, \phi) = (\pi/2, 0)$ is closer to the center specified by $(\xi_*/2, 0)$ than that pair of points. Thus, the minimax location on the Bloch sphere is specified by $(\xi_*/2, 0)$ and φ_{rep} of \mathcal{M}_3 is given by $|\varphi_{rep}\rangle = (\cos \xi_*/4, \sin \xi_*/4)^{\top}$. Figure 2 shows the demand points and minimax location of the facility on the Bloch sphere.



Figure 2. Configuration of demand points in M_3 and the minimax facility location on the Bloch sphere: The red solid points denote demand points and the green solid point denotes the minimax facility location (the representative quantum state), which is the geodesic midpoint between the most distant pair of demand points.

4.1.3. Example 3: \mathcal{M}_4

For \mathcal{M}_4 , we follow Algorithm Step (1) and Step (2) described in the next subsection. After tedious calculation, we find the minimax location, which is specified by $(\xi_*/2, \pi/4)$ $(\xi_*$ is defined as above) and φ_{rep} of \mathcal{M}_4 is given by $|\varphi_{rep}\rangle = (\cos \xi_*/2, e^{i\frac{\pi}{4}} \sin \xi_*/2)^\top$. Figure 3 shows the demand points and the minimax location of the facility on the Bloch sphere.



Figure 3. Configuration of demand points in M_4 and the minimax facility location on the Bloch sphere: The red solid points denote demand points, and the green solid point denotes the minimax facility location (the representative quantum state), which is the center of the circumscribed circle of the triangle whose edges are the demand points (See Algorithm Step (2)).

It may be thought that using the squared Euclidean distance rather than the arc length is unnatural. However, as shown in Lemma 1, The minimax location obtained under the squared Euclidean distance agrees with that obtained under the arc length due to the strict monotonicity of $\alpha \mapsto 4 \sin^2(\alpha/2)$ (See the beginning of Section 4.1 for the squared Euclidean distance). In this sense, *the representative quantum state of a model is invariant*. On the contrary, the least favorable weight for the facility location problem depends on the measure of distance on the sphere; thus, it is not invariant.

4.2. Algorithm to Find Nonrandomized Minimax Location

In operations research, there are several studies on the facility location problem on the unit sphere, where some algorithms to find the minimax facility location are also proposed. Inspired by these studies, we propose a naïve algorithm to find the nonrandomized minimax facility location. Specifically, we consider the facility location problem on a three-dimensional hypersphere in a four-dimensional *real* Euclidean space. This is easily generalized to a general dimension. Basically, a pure state model in a *d*-dimensional Hilbert space is regarded as a subset of a complex projective space $\mathbb{C}P^{d-1}$. A complex projective space $\mathbb{C}P^{d-1}$ is a typical example of a complex manifold but is actually a 2d - 2-dimensional *real* manifold. This fact is sufficient to understand the following argument. (For complex projective space, e.g., see Section 4 in Bengtsson and Życzkowki [29]).

We exclude the possibility of a randomized strategy, although it is sometimes better than any nonrandomized strategy, at least theoretically. (For example, see Section 1.5 in Ferguson [30]). For example, let us consider six demand points, $(\xi, \phi) = (\pi/2 \pm \epsilon, 0)$, $(\pi/2 \pm \epsilon, 2\pi/3)$, $(\pi/2 \pm \epsilon, 4\pi/3)$ on the unit sphere, where ϵ is a small positive constant. Then, a randomized facility location strategy, north pole with probability 1/2 and south pole with probability 1/2, yields the average transportation cost measured by the arc length, $1/2(\pi/2 + \epsilon) + 1/2(\pi/2 - \epsilon) = \pi/2$ for each demand point. Thus, it achieves the minimax location. On the other hand, any nonrandomized strategy yields a higher transportation cost ($>\pi/2$) in the worst case. The algorithm presented below fails in this example. When dim H = 2, all demand points are not covered in a hemisphere if and only if there exists a randomized strategy that is better than any nonrandomized strategy. No simple mathematical condition can assure that nonrandomized minimax is not worse than any randomized strategy when dim H > 2. Thus, we implicitly assume this fact and the existence of a nonrandomized minimax strategy in the Algorithm 1.

Algorithm 1: Find Minimax Facility Location

- (1) Find the most distant pair: Let two distinct demand points *A* and *B* be arbitrary. Then, from Formula (7), find the geodesic midpoint *P*. Then calculate the arc length AP (= BP). Find the maximum arc length R_2 and its center $P_{2,*}$. If every arc length between $P_{2,*}$ and a demand point is not more than R_2 , then $P_{2,*}$ is the minimax location and STOP. If not, then go to Step (2).
- (2) Find the most distant triplet: Let *A*, *B*, and *C* be three arbitrary demand points. Find the center *P* of the circumscribed circle of the triangle $\triangle ABC$ on the hypersphere. Then calculate the arc length AP (= BP = CP). Find the maximum arc length R_3 (> R_2) and its center $P_{3,*}$. If every arc length between $P_{3,*}$ and a demand point is not more than R_3 , then $P_{3,*}$ is the minimax location and STOP. If not, then go to Step (3).
- (3) Find the most distant quadruplet: Let *A*, *B*, *C*, and *D* be four arbitrary demand points. Find the center *P* of the circumscribed sphere of the tetrahedron *ABCD* on the hypersphere. Then, calculate the arc length AP (= BP = CP = DP). Find the maximum arc length $R_4 (>R_3)$ and its center $P_{4,*}$. $P_{4,*}$ is the minimax location and STOP.

Due to monotonicity, we may evaluate the squared Euclidean distance or inner product instead of the arc length between two demand points.

When we generalize the algorithm suitably to $\mathbb{C}P^{d-1}$ (as a 2d - 2-dimensional real manifold), we obtain the algorithm to find the representative quantum state of a model in a *d*-dimensional Hilbert space. In the last subsection, we used the proposed algorithm to find the representative quantum state in some examples. In Section 5, we also demonstrate how to find the representative quantum state following the above algorithm in a specific case.

In the qubit system, the above argument is applied to the mixed states because the Bloch ball is regarded as the hypersphere S_{+}^3 , a hemisphere of a 3-sphere, in a real four-dimensional Euclidean space (e.g., see Section 9.5 in Bengtsson and Życzkowki [29]).

Though the algorithm itself is not our main concern, we briefly mention the efficiency of the algorithm. The computational complexity of each Step (1), (2), and (3) is, respectively, $O(m^2)$, $O(m^3)$, and $O(m^4)$, and clearly it is not efficient. The above problem is reduced to finding the covering sphere for all demand points with the minimum radius (cf. Shiode [19]). Based on this idea, it could be possible to obtain more efficient algorithms even for a continuous model.

The facility location problem on the sphere and finding the representative quantum state of a model (in a two-dimensional Hilbert space) are completely different. It is a

bit surprising that the former problem, which comes from operations research, is helpful for understanding the result in the latter, which comes from quantum physics. What a top manager in a global business really cares about might be essentially the same as a fundamental problem in quantum physics. How does this wonderful connection arise? A mathematician might point out the underlying isomorphism between SU(2) and SO(3) [29]. However, this connection arises mainly from a game-theoretic approach. In other words, the unexpected tie implies the universality and effectiveness of game-theoretic concepts, which are different from information-theoretic concepts such as entropy. This is again emphasized when we introduce the definition of model information in the next section.

5. Quantum Detection Game and Model Information

We have explained how to determine a representative quantum state of a given model. Based on the state, in the present section, we define a new information quantity, model information. Geometrically, this is the maximum radius from the representative quantum state as the center.

The basic strategy to define an information quantity is to introduce a certain imaginary two-person game where one player obtains points according to the information available.

For example, in classical information theory, we consider assigning the ideal code length $-\log p(x)$ to each alphabet x when we know the code distribution p(x). Bob's best score is given by his guessed distribution q(x) and obtains the score $\{-\log q(x)\} - \{-\log p(x)\}$ for each alphabet x. Taking the average with respect to p(x), we obtain the Kullback–Leibler information [23], which is a very fundamental quantity in information theory.

According to Tanaka [20], we consider a quantum detection game as an imaginary two-person game.

5.1. Quantum Detection Game and Definition of Purely Quantum Information

As an example, we introduce a four-dimensional pure state model, M_{FP} and set ϵ between $1/4 < \epsilon < 1$. This consists of the following four-dimensional vectors:

$$\mathcal{M}_{FP} = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4\} \tag{11}$$

$$\varphi_1 = e_1 = \begin{pmatrix} 1\\0\\0\\0 \end{pmatrix}, \varphi_2 = \begin{pmatrix} \sqrt{\epsilon}\\\sqrt{1-\epsilon}\\0\\0 \end{pmatrix}, \varphi_3 = \begin{pmatrix} \sqrt{\epsilon}\\0\\\sqrt{1-\epsilon}\\0 \end{pmatrix}, \varphi_4 = \begin{pmatrix} \sqrt{\epsilon}\\0\\0\\\sqrt{1-\epsilon} \end{pmatrix}.$$
(12)

While dim $\mathcal{H} = 4$, it is enough to consider each vector in a *real* four-dimensional vector space.

Let us explain the quantum detection game under the model \mathcal{M}_{FP} . First, Alice picks one pure state from the model (i.e., $\varphi_1, \ldots, \varphi_4$) and then sends it to Bob. Bob knows only the candidate pure state sets and the model, and prepares a two-outcome measurement in the form $\{|\varphi\rangle\langle\varphi|, I - |\varphi\rangle\langle\varphi|\}$, where *I* is the identity operator and φ is the unit vector. We call φ a *detector* or a *detector state*. Bob's purpose is not to guess the number that Alice has chosen but to obtain "detection" with a higher probability.

The detection rate for the chosen state $|\varphi_j\rangle\langle\varphi_j|$ is given by $\text{Tr}\{|\varphi_j\rangle\langle\varphi_j||\varphi\rangle\langle\varphi_j|\} = |\langle\varphi|\varphi_j\rangle|^2$ when Alice sends φ_j to Bob and Bob prepares φ as a detector. (Tr denotes the matrix trace and $|a\rangle\langle b|$ is regarded as a matrix). As a game, Alice aims at making Bob's detection rate smaller by choosing $\varphi_1, \ldots, \varphi_4$ with a certain probability. In contrast, Bob aims at making the detection rate larger by preparing his detector φ based on the knowledge of the model. Later, we will evaluate the information of the model \mathcal{M}_{FP} .

Now we go back to the general situation and explain the details. First, we seek the minimum detection rate for Bob among all possible models. Suppose that Alice picks among the whole pure states in a completely random way (i.e., with respect to the Haar measure). This is the worst case for Bob. When Bob is allowed to adopt a randomized

strategy, the detection rate is 1/d (*d* is the dimension of the Hilbert space). If the model consists of the orthonormal basis, then again the detection rate is 1/d. It is the minimum detection rate.

Next, suppose that Alice has a certain tendency for choosing the pure state, which is also described by the model \mathcal{M} and Bob knows this for some reason. Although we do not care about the origin of such models, there are various situations where they apply in quantum science and technology. For example, in the bipartite system $\mathbb{C}^2 \otimes \mathbb{C}^2$ without interaction, a pure state arises as a product state like $|\varphi\rangle|\varphi'\rangle$. Then an entangled state such as $\alpha|00\rangle + \beta|11\rangle$ is not expected. In quantum computation, the output qubit state under the unitary gate, which has some rotation error, would be $e^{i(\epsilon+\pi/2)Y}|\varphi\rangle$. Then, Bob could obtain a detection rate larger than 1/d based on the information of the model. Following this idea, we propose one information quantity for a model below.

A detailed explanation of the quantum detection game and useful results are described in the author's previous work [20]. Below, we only present some of the results in a formal way, which is necessary to define the information quantity. Those definitions hold in an infinite-dimensional Hilbert space.

First, we define the Bayes mixture in a slightly formal way. Let $\mathcal{M} = \{\psi_{\theta} : \theta \in \Theta\}$ be a model, (see Section 3) and π be a probability distribution on the parameter space Θ . Then, the Bayes mixture ρ_{π} is defined as

$$\rho_{\pi} = \int_{\Theta} |\psi_{\theta}\rangle \langle \psi_{\theta} | \pi(d\theta), \tag{13}$$

In the context of Bayesian statistics [31–33], we call π a *prior distribution* or briefly a *prior*. For a discrete model, the above integral is replaced with a finite sum $\sum_j \pi_j |\psi_j\rangle \langle \psi_j|$. Then, when Alice sends $|\psi_j\rangle$ to Bob with probability π_j , it is equivalent to sending ρ_{π} to Bob in the quantum detection game.

Finally, we have come to our main theme: to define the information quantity of a model \mathcal{M} .

Definition 2. Let \mathcal{M} be a model in a d-dimensional Hilbert space. Then, we define the purely quantum information (PQI) of a model \mathcal{M} as

$$J(\mathcal{M}) = \inf_{\pi} \|\rho_{\pi}\|_{\infty} - 1/d.$$

For calibration, we subtract the lower bound 1/d, and thus $J(\mathcal{M}) \ge 0$. When Bob knows that the quantum state Alice prepares is among \mathcal{M} , we interpret this as Bob obtaining $J(\mathcal{M})$. As shown in Section 5.3, the above infimum is related to the value of the quantum detection game (possible maximum score) through the minimax theorem [20].

Let us rewrite $J(\mathcal{M})$ in a slightly simpler form. For a discrete model, there exists a prior distribution that achieves the infimum of $\|\rho_{\pi}\|_{\infty}$. Then, we call the prior a *least favorable prior* (LFP). LFP is one of the technical terms in statistical decision theory or game theory (see, e.g., Section 1.7, p. 39 in Ferguson [30]). Using the least favorable prior π_{LF} , PQI is defined by

$$J(\mathcal{M}) = \|\rho_{\pi_{LF}}\|_{\infty} - 1/d.$$
 (14)

Even if LFP is not uniquely determined, $\|\rho_{\pi_{LF}}\|_{\infty}$ remains the same [20]. As some readers may recognize, the LFP completely agrees with the least favorable weight in Section 4.

5.2. Basic Properties of PQI

From the form (14), we obtain some properties of PQI easily. First, by definition, J(M) is independent of a basis. In other words, both $M = \{\psi_j\}$ and $M' = \{U\psi_j\}$, where *U* is a unitary operator, yield the same PQI. Second, clearly the following holds.

Lemma 4. Let \mathcal{M} be a model in a d-dimensional Hilbert space. The following conditions are equivalent.

- (i) $J(\mathcal{M}) = 0.$
- (*ii*) $\rho_{\pi_{LF}} = \frac{1}{d}I$ for every LFP.

When J(M) = 0, Alice can send the completely mixed state effectively and then Bob obtains no information from the model M to achieve a higher detection rate than 1/d. Geometrically speaking, such a model fully spreads with no specific direction.

In contrast, when $J(\mathcal{M}) > 0$, a certain bias exists and it prevents Alice from preparing the completely mixed state. Thus, Bob benefits after knowing the model. If \mathcal{M} satisfies the full-rank condition, then there exists a prior π such that the Bayes mixture $\rho_{\pi} > 0$ (A > 0 denotes the positive definiteness of a Hermitian matrix A). If \mathcal{M} does not satisfy the full-rank condition, we have a d'-dimensional subspace where \mathcal{M} (restricted to the subspace) satisfies the full-rank condition. Since $\inf_{\pi} \|\rho_{\pi}\|_{\infty} \ge 1/d' > 1/d$, we have the lower bound of PQI, $J(\mathcal{M}) \ge 1/d' - 1/d$.

We mention the relation between PQI and the von Neumann entropy $S(\rho)$. (Recall that the von Neumann entropy is defined by $S(\rho) = -\text{Tr}\rho \log \rho$). It is easily shown that $J(\mathcal{M}) = 0$ if and only if $S(\rho_{\pi_{LF}}) = \log d$. The worst case for Bob also corresponds to the maximum entropy state. As we shall see in Section 6, our formulation is instead related to the minimum entropy.

Next, we consider how to calculate the PQI of a given model. If the model has a certain symmetry, then we obtain the LFP analytically and directly calculate $\|\rho_{\pi_{LF}}\|_{\infty}$. On the other hand, due to the minimax theorem in the author's previous work [20], the infimum of the operator norm of a Bayes mixture, $\inf_{\pi} \|\rho_{\pi}\|_{\infty}$ is easily calculated by finding the representative quantum state of the model, which is defined in Section 3. Thus, we are able to calculate the PQI of a given model by finding the minimax point (the representative quantum state of the model), and to do so, we utilize the algorithm shown in Section 4 in order to find the minimax point in the facility location problem on the unit sphere. We present the above procedure explicitly in Section 5.4 in detail.

The mathematical structure is quite similar to the calculation of channel capacity in classical information theory [34,35]. However, we emphasize that even in a formal analogue, we do not introduce any entropic quantity or any concept from information theory to define the above PQI. What we have used is an imaginary two-person game called the quantum detection game and some basic rules in quantum physics. Taking into account many works in quantum information theory [36], it is a bit surprising to develop purely quantum information *without referring to any classical concepts in information theory* [37–40].

We also note that PQI is completely different from other kind of information quantity such as the Fisher information [7,8]. For a parametric model of quantum states, $\rho(\theta)$, differentiable with respect to the parameter θ , Fisher information evaluates the change of quantum states, $\rho(\theta + \Delta \theta) - \rho(\theta)$. It is related to the distinguishability between two close quantum states $\rho(\theta)$ and $\rho(\theta + \Delta \theta)$ from observation *after* performing some measurements. Let us take a specific example to see the difference. Suppose that we have a continuous one-parameter model $\mathcal{M}_{rot} = \{|\varphi(s)\rangle = (\cos s)|0\rangle + (\sin s)|1\rangle : 0 \le s \le \pi/4\}$. Although quantum Fisher information has been defined in various ways as an extension of classical Fisher information, it is not defined for a discrete model such as $\mathcal{M}_2 = \{|0\rangle, |+\rangle\}$. Indeed, for \mathcal{M}_2 , we only consider distinguishing between two possible states (quantum state discrimination), while for \mathcal{M}_{rot} , we have to consider parameter estimation (quantum state estimation), and the estimation error is bounded by SLD Fisher information [4–6]. However, PQI yields the same value for both \mathcal{M}_2 and \mathcal{M}_{rot} .

5.3. Basic Formula for PQI Calculation

We provide several examples to show how we calculate the PQI of a model below. We give the following formula, which connects the representative quantum state and PQI. The formula is obtained by the minimax theorem (10).

$$- \|\rho_{\pi_{LF}}\|_{\infty} = \min_{\varphi} \max_{\psi \in \mathcal{M}} d_F(\varphi, \psi)$$
$$= \max_{\psi \in \mathcal{M}} d_F(\varphi_{rep}, \psi).$$
(15)

Or equivalently, we have the formula

1

$$\|\rho_{\pi_{LF}}\|_{\infty} = \max_{\psi \in \mathcal{M}} |\langle \varphi_{rep} | \psi \rangle|^2.$$
(16)

Using the above formula and result in Section 4, we obtain the PQI of M_2 , M_3 , and M_4 (For the definition, see Section 2.1).

First, we consider the PQI of \mathcal{M}_2 . We already know the representative quantum state of \mathcal{M}_2 from Section 4. Thus, using the Formula (16), we obtain $\|\rho_{\pi_{LF}}\|_{\infty} = |\langle \varphi_{rep}|0\rangle|^2 = \frac{2+\sqrt{2}}{4}$ and $J(\mathcal{M}_2) = \frac{\sqrt{2}}{4}$. In the same way, for \mathcal{M}_3 , using the Formula (16), we obtain $\|\rho_{\pi_{LF}}\|_{\infty} = \frac{3+\sqrt{3}}{6}$, $J(\mathcal{M}_3) = \frac{\sqrt{3}}{6}$. Straightforward calculation also yields $J(\mathcal{M}_4) = \frac{\sqrt{3}}{6}$, which coincidently is equal to $J(\mathcal{M}_3)$.

5.4. Example of PQI Calculation: \mathcal{M}_{FP}

Next, as a more nontrivial example, we calculate the PQI of the model M_{FP} introduced in Section 5.1. First, following the algorithm in Section 4, let us find the minimax point (the representative quantum state of the model M_{FP}).

In Step (1), we find the most distant pair. We mainly focus on the inner product between two vectors instead of the geodesic distance between them. Then, the most distant pair corresponds to those for which the inner product is the closest to zero. Since $|\langle \varphi_1 | \varphi_j \rangle|^2 = \epsilon$, $|\langle \varphi_j | \varphi_k \rangle|^2 = \epsilon^2$, (j, k = 2, 3, 4), and $0 < \epsilon^2 < \epsilon < 1$, the most distant pair is $\{\varphi_2, \varphi_3\}, \{\varphi_2, \varphi_4\}, \text{ and } \{\varphi_3, \varphi_4\}.$

Using the Formula (7) in Lemma 2, we obtain the geodesic midpoint between φ_2 and φ_3 , which is denoted by M_1 and $\varphi_{M_1} = \frac{1}{\sqrt{2(1+\epsilon)}} (2\sqrt{\epsilon}, \sqrt{1-\epsilon}, \sqrt{1-\epsilon}, 0)^\top$. Comparing the inner products, it is easily seen that φ_1 is located at a point more distant from φ_{M_1} than φ_2 and φ_3 . Thus, we go to Step (2) in the algorithm. Note that in the model \mathcal{M}_{FP} , all inner products are real and positive.

In Step (2), we find the most distant triplet. In our model, it is enough to consider the circumscribed hypercircle in a *real* four-dimensional Euclidean space. Due to the symmetry, we only check two triangles, Δ_{234} and Δ_{123} whose vertices are { φ_2 , φ_3 , φ_4 } and { φ_1 , φ_2 , φ_3 }, respectively.

First, let *Q* be the center of the circumscribed hypercircle of the triangle Δ_{234} . (Each edge is a geodesic on the sphere). Generally, the point *Q* is not uniquely determined. However, by imposing the condition that *Q* is on the three-dimensional real subspace $L_{234} = \text{span}_{\mathbf{R}}\{\varphi_2, \varphi_3, \varphi_4\}$, the point *Q* is uniquely determined as the point achieving the minimum distance from each vertex (radius of the circumscribed hypercircle). The condition is equivalent to an orthogonality condition, i.e., $\psi \in L_{234} \iff \langle \psi | \varphi_L \rangle = 0$, where $|\varphi_L\rangle = (\sqrt{1-\epsilon}, -\sqrt{\epsilon}, -\sqrt{\epsilon}, -\sqrt{\epsilon})^\top$.

Now let $\varphi_Q = (x, y, z, w)^\top$, x, y, z, w > 0 be a vector corresponding to Q. Then it satisfies $\langle \varphi_Q | \varphi_2 \rangle = \langle \varphi_Q | \varphi_3 \rangle = \langle \varphi_Q | \varphi_4 \rangle$, $\langle \varphi_Q | \varphi_L \rangle = 0$, and $\| \varphi_Q \|_2^2 = 1$, where $\| \cdot \|_2$ denotes the Euclid norm. We obtain the solution $\varphi_Q = \frac{1}{\sqrt{3(2\epsilon+1)}} \left(3\sqrt{\epsilon}, \sqrt{1-\epsilon}, \sqrt{1-\epsilon}, \sqrt{1-\epsilon} \right)^\top$.

Next, we investigate the other circumscribed hypercircle of the triangle Δ_{123} . In a similar way, we define the point *R* for Δ_{123} . Then, the state vector corresponding to *R* is

given by
$$\varphi_R = \left(\sqrt{\frac{1+\sqrt{\epsilon}}{3-\sqrt{\epsilon}}}, \sqrt{\frac{1-\sqrt{\epsilon}}{3-\sqrt{\epsilon}}}, \sqrt{\frac{1-\sqrt{\epsilon}}{3-\sqrt{\epsilon}}}, 0\right)^{'}$$
.

Let each radius of the circumscribed hypercircle be r_{123} and r_{234} . Then $r_{123}^2 = d_F(\varphi_R, \varphi_1) = \frac{2(1-\sqrt{\epsilon})}{3-\sqrt{\epsilon}}$, and $r_{234}^2 = d_F(\varphi_Q, \varphi_2) = \frac{2(1-\epsilon)}{3}$. Thus, $r_{234}^2 > r_{123}^2$, the most distant triplet is $\{\varphi_2, \varphi_3, \varphi_4\}$.

Finally, we check whether the circumscribed hypercircle with center φ_Q and radius r_{234} includes the other point φ_1 . (If not, we go to Step (3) in the algorithm) Since we assume that $(1 >)\epsilon > 1/4$, $\langle \varphi_1 | \varphi_Q \rangle = \sqrt{\frac{3\epsilon}{1+2\epsilon}} > \langle \varphi_2 | \varphi_Q \rangle = \langle \varphi_3 | \varphi_Q \rangle = \langle \varphi_4 | \varphi_Q \rangle = \sqrt{\frac{1+2\epsilon}{3}}$ holds, this implies that φ_1 is closer to the point φ_Q than the other three points. Thus, the Algorithm stops and φ_Q is the minimax location.

Using φ_Q , we obtain

$$egin{aligned} V_L &:= \sup_{egin{subarray}{c} \psi, \|\psi\|_2 = 1} \inf_{1 \leq j \leq 4} |\langle arphi_j |\psi
angle|^2 \ &= rac{1+2\epsilon}{3}. \end{aligned}$$

Due to Equation (16), V_L agrees with the infimum of the detection rate $\|\rho_{\pi_{LF}}\|_{\infty}$, and we obtain PQI $J(\mathcal{M}_{FP}) = \frac{1+8\epsilon}{12}$.

5.5. PQI Calculation from LFP

Now let us find the LFP in this model. Since the model M_{FP} has a certain symmetry, we obtain it directly.

First, let the support of π be { $\varphi_2, \varphi_3, \varphi_4$ }, that is, $\pi_2 + \pi_3 + \pi_4 = 1$ and $\pi_1 = 0$, $\pi_j > 0, j = 2, 3, 4$. Then we obtain one of the LFPs, which is given by the uniform distribution, $\pi_2 = \pi_3 = \pi_4 = 1/3$. To see this, we use the following two facts. First, for every permutation σ (e.g., $\sigma(2) = 4, \sigma(3) = 3, \sigma(4) = 2$),

$$\left\|\sum_{j=2}^{4} \pi_{\sigma(j)} |\varphi_{j}\rangle \langle \varphi_{j}|\right\|_{\infty} = \left\|\sum_{j=2}^{4} \pi_{j} |\varphi_{j}\rangle \langle \varphi_{j}|\right\|_{\infty}$$

holds. (To see this, construct the unitary operator U_{σ} , which is the group homomorphism $\sigma \mapsto U_{\sigma}$). Second, the average for every permutation is given by

$$\frac{1}{3!}\sum_{\sigma\in\mathfrak{S}_{234}}\sum_{j}\pi_{\sigma(j)}|\varphi_{j}\rangle\langle\varphi_{j}|=\frac{1}{3}\sum_{j}|\varphi_{j}\rangle\langle\varphi_{j}|,$$

where \mathfrak{S}_{234} is the permutation group acting on the set $\{2, 3, 4\}$. Thus, for every π ,

$$\begin{split} \left\| \frac{1}{3} \sum_{j} |\varphi_{j}\rangle \langle \varphi_{j}| \right\|_{\infty} &= \left\| \frac{1}{3!} \sum_{\sigma \in \mathfrak{S}_{234}} \sum_{j} \pi_{\sigma(j)} |\varphi_{j}\rangle \langle \varphi_{j}| \right\|_{\infty} \\ &\leq \frac{1}{3!} \sum_{\sigma \in \mathfrak{S}_{234}} \left\| \sum_{j} \pi_{\sigma(j)} |\varphi_{j}\rangle \langle \varphi_{j}| \right\|_{\infty} \\ &= \left\| \sum_{j} \pi_{j} |\varphi_{j}\rangle \langle \varphi_{j}| \right\|_{\infty} \end{split}$$

holds. Thus, the uniform distribution is the LFP when $\pi_1 = 0$.

Then, we relax the condition $\pi_1 = 0$. We set the uniformly mixed state as

$$\rho_{\pi_*} = \frac{1}{3} |\varphi_2\rangle \langle \varphi_2| + \frac{1}{3} |\varphi_3\rangle \langle \varphi_3| + \frac{1}{3} |\varphi_4\rangle \langle \varphi_4|.$$
(17)

Then, the operator norm is given by

$$\|\rho_{\pi_*}\|_{\infty} = \frac{1+2\epsilon}{3}.$$
(18)

From direct but very tedious calculation, we can see that the above achieves the infimum even if we permit $\pi_1 > 0$. Setting $\rho_{\pi} := \sum_{j=1}^4 \pi_j |\varphi_j\rangle \langle \varphi_j |$, we obtain

$$\begin{split} \inf_{\pi} \|\rho_{\pi}\|_{\infty} &= \inf_{0 \leq p \leq 1} \|p\rho_{\pi_{*}} + (1-p)|\varphi_{1}\rangle \langle \varphi_{1}| \parallel_{\infty} \\ &= \frac{1+2\epsilon}{3}. \end{split}$$

Therefore, we obtain the LFP as

$$\pi_{*,1} = 0, \ \pi_{*,2} = \pi_{*,3} = \pi_{*,4} = 1/3.$$
 (19)

Even when we do not find the representative quantum state directly, we can construct it from the LFP in the following way. Since the Bayes mixture with respect to LFP is given by Equation (17), we find the first eigenvector ψ_{π_*} with the maximum eigenvalue ρ_{π_*} (no de-

generacy), which is given by
$$\psi_{\pi_*} = \left(\sqrt{\frac{3\epsilon}{2\epsilon+1}}, \sqrt{\frac{1-\epsilon}{3(2\epsilon+1)}}, \sqrt{\frac{1-\epsilon}{3(2\epsilon+1)}}, \sqrt{\frac{1-\epsilon}{3(2\epsilon+1)}}\right)^{-1}$$

Actually, this agrees with the representative quantum state, φ_Q , in Section 5.4.

Through the minimax theorem [20], we can also directly show that the norm (18) achieves the minimum. To see this, we introduce the following inequality:

$$\begin{split} V_U &:= \inf_{\pi} \sup_{\psi, \|\psi\|_2 = 1} \sum_{j=1}^4 \pi_j |\langle \varphi_j |\psi \rangle|^2 \\ &= \inf_{\pi} \|\rho_{\pi}\|_{\infty} \\ &\leq \frac{1+2\epsilon}{3}. \end{split}$$

Since $V_U \ge V_L$ (for details, see the author's previous work [20]), we obtain $\frac{1+2\epsilon}{3} \ge V_U \ge V_L = \frac{1+2\epsilon}{3}$, which implies $V_U = V_L$. Thus, one LFP is given by (19). The above argument does not exclude another possibility for the LFP with $\pi_1 > 0$.

5.6. Difference from Maximization of von Neumann Entropy

In our formulation, PQI has no direct relation to any entropic concept. Since some readers may expect a certain relationship, let us see what happens if we formally adopt the von Neumann entropy to obtain the LFP in the last example. We consider the maximization of $S(\rho_{\pi})$ over the prior π . The concavity of $S(\rho)$ yields

$$\begin{aligned} \max_{\pi} S(\rho_{\pi}) &= \max_{0 \leq \lambda \leq 1} S(\lambda |\varphi_{1}\rangle \langle \varphi_{1}| + (1 - \lambda)\rho_{\pi_{*}}) \\ &\geq \max_{0 \leq \lambda \leq 1} \{\lambda S(|\varphi_{1}\rangle \langle \varphi_{1}|) + (1 - \lambda)S(\rho_{\pi_{*}})\} \\ &= S(\rho_{\pi_{*}}), \end{aligned}$$

where π_* denotes LFP (19). For some ϵ , we numerically find a positive λ_* achieving the maximum of $S(\lambda | \varphi_1 \rangle \langle \varphi_1 | + (1 - \lambda) \rho_{\pi_*})$. Thus, a positive weight for $|\varphi_1 \rangle \langle \varphi_1 |$ could appear under the maximization, which is clearly different from our result.

While the LFP π_* , which is obtained by minimization of $\|\rho_{\pi}\|_{\infty}$, yields the minimax solution in the quantum detection game, the maximizer, say, π_{ent} , is meaningless, at least in this example. Indeed, $\|\rho_{\pi_{ent}}\|_{\infty} > \|\rho_{\pi_*}\|_{\infty}$, which implies that $\rho_{\pi_{ent}}$ is more informative than ρ_{π_*} , and the prior π_{ent} is not the least favorable to Bob anymore.

We have carefully treated the information or uncertainty of a nonorthogonal pure state model and excluded classical fluctuation. As a consequence, the remaining uncertainty is not evaluated by the usual entropy anymore. For a nonorthogonal pure state model, the von Neumann entropy as a measure of information lacks theoretical justification. At least in the quantum detection game, the method based on the von Neumann entropy is a mere formal extension. It makes sense only for a model that consists of orthogonal pure states (see Section 2).

However, there are many variants of entropy [37–41] both in classical and quantum information theory. We discuss a certain relationship between our information quantity and the minimum entropy in the next section.

6. Discussion: Relation to Entropy

In the previous section, we introduced an information quantity for a pure state model called PQI. Under the full-rank condition, any classical model consists of an orthonormal basis. Then PQI of the model necessarily vanishes.

We emphasize that PQI is literally something *purely quantum* since we have not formally extended something in classical information theory. It is the information quantity completely independent of the concept of entropy, which does make sense in classical information theory. Thus, a natural question arises, i.e., what kind of relationship do the entropy and PQI have? Actually, PQI is related to the minimum entropy instead of the von Neumann entropy, as we will discuss below.

6.1. Jaynes Principle and Distinguishability

First, we briefly review the concept of entropy and the Jaynes principle [42,43].

Suppose that we are given the set of the alphabet. Then our lack of knowledge on the set is evaluated by Shannon entropy through a probability distribution $\{p_i\}$ satisfying $p_1 + \cdots + p_d = 1$, $p_i \ge 0$, $i = 1, \ldots, d$. (Recall that classical Shannon entropy is defined as $S_{cl}(p) = \sum_i -p_i \log p_i$). The larger the entropy becomes, the larger the uncertainty we have. We have minimum information as interpreted as the maximum entropy state, that is, $p_i = 1/d$ and thus

$$\max_{p} S_{cl}(p) = \log d$$

holds. The central idea also provides the theoretical foundation for maximum entropy methods in data processing [44,45].

The underlying concept is distinguishability. In classical information theory, distinguishability holds trivially. In quantum theory, it is represented by the orthogonality of two quantum states. When pure states corresponding to alphabets, say, $\{1, 2, ..., d\}$, are orthogonal to each other, every result in classical information theory is extended in a straightforward manner.

In statistical physics, a physical state of an ensemble is estimated through entropy maximization when we have no knowledge of the system. This way of thinking is called the *Jaynes principle* [42,43], and it is fundamental to statistical physics. For example, for a given set of eigenstates of a Hamiltonian, say, $|E_0\rangle$, $|E_1\rangle$,..., with some conditions, we obtain a canonical ensemble by using the principle.

In quantum physics, we are able to consider the maximization of the von Neumann entropy of the density matrix $\rho_{\pi} = \sum_{j} \pi_{j} |\psi_{j}\rangle \langle \psi_{j}|$ (Bayes mixture) for orthogonal vectors $\{\psi_{1}, \ldots, \psi_{d}\}$. Since $S(\rho_{\pi}) = S_{cl}(\pi)$, this maximization completely reduces to the classical case. Then the maximizer is the completely mixed state, i.e., $\frac{1}{d}I$, which corresponds to the

uniform distribution. Formally, additional constraints also yield a quantum exponential family [46], which is the quantum analogue of the classical exponential family [47,48].

However, we have no solid criterion such as the Jaynes principle for a nonorthogonal pure states model. For example, a qubit $|0\rangle$ processed under one unitary operation, which is assumed to be among U_1, U_2, U_3 . $(|\psi_j\rangle = U_j|0\rangle, j = 1, 2, 3)$. In a sense, it is a simplified rotation error model (e.g., Kueng et al. [2]). In our formulation, a model $\mathcal{M}_U = \{\psi_1, \psi_2, \psi_3\}$ is given. Suppose that we have no information or knowledge on which unitary gate processed the qubit. Then, how do we describe the quantum bit?

Mathematically, it is possible to extend the maximum entropy criterion to the noncommutative case. Then we consider the maximization,

$$\sup_{\pi} S(\rho_{\pi}) \tag{20}$$

over the prior π . Is this kind of formal extension enough in quantum information theory? There are many quantities such as Rényi's entropy [37–41] in both classical and quantum information theory. Is there another possibility to consider such quantities?

Every quantum state in the model M_U is not orthogonal anymore; thus, they are not distinguishable, which is completely different from the set of the alphabet. In spite of this, do we seek some justification for the maximization of the entropy from classical information theory?

In our formulation, the above formal argument breaks down. First, for the model \mathcal{M}_U , we describe the system by the representative quantum state $\varphi_{rep}(\mathcal{M}_U)$, which is completely independent of the von Neumann entropy. Second, in the quantum detection game between Alice and Bob, we see that the von Neumann entropy is useless in a specific example (Section 5.6). Rather, we consider the least favorable case or the minimization of the detection rate, $\inf_{\pi} \|\rho_{\pi}\|_{\infty}$, which is contrastive to the maximization of entropy (20).

If we seek a purely quantum counterpart of the Jaynes principle, then minimization of $\|\rho_{\pi}\|_{\infty}$ would be promising. Luckily, due to monotonicity, the minimization is equivalent to the maximization of $-\log \|\rho_{\pi}\|_{\infty}$, where the function $-\log \|\rho\|_{\infty}$ is known as the *minimum entropy* of ρ . Some of its properties are similar to those of the von Neumann entropy and others are not. In the next subsection, we review basic properties of the minimum entropy.

6.2. Properties of the Minimum Entropy

In the present subsection, we briefly review basic properties of the minimum entropy and then give another definition for purely quantum information. The minimum entropy of the density matrix ρ is defined by $T(\rho) := -\log \|\rho\|_{\infty}$, which is a special case of quantum Rényi entropy.

Quantum Rényi entropy has a real parameter α and is defined by

$$S_{\alpha}(\rho) = \frac{\log \mathrm{Tr} \rho^{\alpha}}{1-\alpha}$$

for fixed α (see, e.g., p. 117 in Ohya and Petz [41]). When $\alpha \to \infty$, we obtain the minimum entropy.

The minimum entropy $T(\rho)$ inherits some common properties of quantum Rényi entropy. For example, additivity $T(\rho \otimes \sigma) = T(\rho) + T(\sigma)$ holds. For every pure state, *T* equals to zero.

However, the concavity does not necessarily hold. Concavity of entropy means that a probability mixture of quantum states increases the uncertainty of the whole system.

This negative property is not due to noncommutativity. To see this, let us take two commutative density matrices,

$$\rho = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \ \sigma = \begin{pmatrix} \mu & 0 \\ 0 & 1 - \mu \end{pmatrix},$$

where $\frac{1}{2} \le \mu < 1$. Then, since $\left\| \frac{\rho + \sigma}{2} \right\|_{\infty} = \frac{1 + \mu}{2}$,

$$\frac{1}{2}T(\rho) + \frac{1}{2}T(\sigma) - T\left(\frac{\rho+\sigma}{2}\right) = \log\left(\frac{1+\mu}{2\sqrt{\mu}}\right) > 0$$

Thus, convexity rather than concavity holds in the above example.

Since minimum entropy is based on the operator norm, we obtain a sufficient condition for convexity easily.

Lemma 5. Let two density matrices ρ and σ exist where $\|\rho\|_{\infty} \neq \|\sigma\|_{\infty}$. Suppose that

$$\|\lambda\rho + (1-\lambda)\sigma\|_{\infty} = \lambda\|\rho\|_{\infty} + (1-\lambda)\|\sigma\|_{\infty}$$

holds for every λ *,* $0 < \lambda < 1$ *. Then*

$$T(\lambda \rho + (1 - \lambda)\sigma) < \lambda T(\rho) + (1 - \lambda)T(\sigma)$$

holds.

Proof. Since $-\log(x)$ is convex,

$$T(\lambda\rho + (1-\lambda)\sigma) = -\log \|\lambda\rho + (1-\lambda)\sigma\|_{\infty}$$

= $-\log\{\lambda\|\rho\|_{\infty} + (1-\lambda)\|\sigma\|_{\infty}\}$: Assump.
 $<\lambda(-\log\|\rho\|_{\infty}) + (1-\lambda)(-\log\|\sigma\|_{\infty})$
= $\lambda T(\rho) + (1-\lambda)T(\sigma)$

holds. \Box

To see the above lemma, when we introduce $T(\rho)$ as a variant of entropy over the whole density operators, its theoretical significance seems very weak.

However, when we consider PQI in a pure state model, the situation changes drastically. For the pure state family, concavity of the minimum entropy necessarily holds in the following sense.

Lemma 6. Let a set of pure states be given, say \mathcal{M} . Then concavity holds, restricted to the model.

Proof. Choose a finite set of pure states from \mathcal{M} , say ρ_1, \ldots, ρ_k . Then

$$T\left(\sum_{j=1}^{k} \lambda_j \rho_j\right) \ge 0 = \sum_{j=1}^{k} \lambda_j T(\rho_j)$$

holds. \Box

Other properties of minimum entropy are usually shown in the context of quantum Rényi entropy (see, e.g., Hu and Ye [22] (Section III, p. 4), Dam and Hayden [21]).

Observing the above, we provide another possible definition of purely quantum information through $T(\rho)$. In the quantum detection game, finding the LFP is equivalent to maximizing the minimum entropy $T(\rho_{\pi})$ rather than the von Neumann entropy $S(\rho_{\pi})$. To consider the logarithm of the detection rate, we obtain another definition of purely quantum information such as

$$J'(\mathcal{M}) = \log d - \sup_{\pi} T(\rho_{\pi}) \ (\geq 0).$$

By definition, $J'(\mathcal{M})$ vanishes if the model has orthogonal pure states with the fullrank condition, where the classical case is included. From Lemmas 5 and 6, we must be careful to treat minimum entropy $T(\rho)$ in the definition of $J'(\mathcal{M})$. At least, it should not be considered over the set of the whole density matrices. As a consequence, a comparison of the two definitions, $J(\mathcal{M})$ and $J'(\mathcal{M})$, also should be performed carefully and would require a deeper understanding of the model information, which will be a topic for future research.

Finally, we make two comments. First, our definition of the model information yields one operational meaning for the minimum entropy. It is apart from the usual extension of entropic concepts in classical information theory. Rather, it comes from an imaginary design of the quantum detector and facility location problem on the unit sphere in a complex projective space. Second, we expect that the purely quantum version of the Jaynes principle is established based on the minimum entropy. (For some related works on maximum entropy methods, see the reference [49]). It might be possible to develop data processing methods and some dynamics based on the new principle.

7. Infinite-Dimensional Hilbert Space

Thus far, we have considered the PQI of a model only in a finite-dimensional Hilbert space. While our definition of PQI applies to infinite-dimensional Hilbert space, technical difficulties seem to arise due to a parametric family of functions. In this section, we only skim them in a specific example.

Let $L^2(\mathbf{R})$ denote the set of square integrable complex functions over \mathbf{R} and g be a known continuous function in $L^2(\mathbf{R})$ satisfying $||g||_2^2 = \int |g(x)|^2 dx = 1$. Let us consider the quantum statistical model $\mathcal{M}_{\infty} = \{g(x - \theta) : \theta \in \mathbf{R}\}$ describing a wavefunction with a single parameter.

Parameter estimation of the shift parameter θ has been theoretically investigated [3]. If we replace the wavefunction *g* with a probability density function such as the Gaussian density, the estimation problem for the shift parameter θ is called that for the location parameter and is very common in classical statistics [50].

Before evaluating the PQI of the model \mathcal{M}_{∞} , first let us formally consider quantum state estimation with no observation. It is seen that the worst-case error equals one for every estimation.

Lemma 7. Let
$$g_{\theta}(x) = g(x - \theta)$$
 and $f \in L^{2}(\mathbf{R})$ with $||g_{\theta}||_{2} = ||f||_{2} = 1$. Then

$$\inf_{\theta \in \mathbf{R}} |\langle g_{\theta} | f \rangle| = 0, \tag{21}$$

$$\sup_{\theta \in \mathbf{R}} \left\{ 1 - \left| \langle g_{\theta} | f \rangle \right|^2 \right\} = 1$$
(22)

holds.

For proof, see the Appendix A.1. The above lemma says that every quantum state in $L^2(\mathbf{R})$ would be a minimax location on " $\mathbb{C}P^{\infty ''}$.

Since the parameter space $\Theta = \mathbf{R}$ is noncompact, the minimax theorem [20] does not hold generally. However, we directly show that the Formula (16) holds in this specific example, that is,

$$\inf_{\pi} \|
ho_{\pi} \|_{\infty} = 0 = \sup_{f \in L^2(\mathbf{R})} \left\{ \inf_{ heta \in \mathbf{R}} |\langle g_{ heta} | f
angle |^2
ight\}.$$

The first equality holds due to the following lemma. Because of technical difficulties, we give a proof in the Appendix A.2.

Lemma 8. Let $g_{\theta}(x) = g(x - \theta)$ with $||g_{\theta}||_2 = 1$ and ϵ be an arbitrary positive constant. Then there exists a finite set $\{\theta_1, \ldots, \theta_n\}$ and the uniform prior π_n over the set such that

$$egin{aligned} \|
ho_{\pi_n}\|_\infty &\leq rac{1}{n} + \epsilon, \
ho_{\pi_n} &= rac{1}{n}\sum_{j=1}^n |g_{ heta_j}
angle \langle g_{ heta_j}| \end{aligned}$$

Thus, a formal definition of PQI shows that $J(\mathcal{M}_{\infty}) = \inf_{\pi} \|\rho_{\pi}\|_{\infty} = 0$. We can interpret the result as follows. Even if Bob knows that the quantum state is in the model \mathcal{M}_{∞} or that the quantum system is described by a wavefunction $g(x - \theta)$, he obtains no information, which gives Bob an advantage over Alice in the quantum detection game.

We do not obtain the conditions where PQI is positive with explicit examples. Even if the Formula (16) holds under some conditions, calculation of PQI would become drastically different. A detailed investigation is left for future study.

8. Concluding Remarks

We have defined one information quantity called purely quantum information (PQI) not for a pure state itself, but for a parametric model of pure states. While PQI evaluates the size of a pure state model, it necessarily vanishes in classical cases by definition. We call the center of the model the representative quantum state, and PQI is determined by the maximum length from the center to each quantum state in the model.

Finally, we give the answer to the problem presented in the beginning of the article. Let $\psi_{\theta} = e^{-i\theta_2 Z} e^{-i\theta_1 Y} |0\rangle$ and calculate PQI for two models:

$$\mathcal{M}_5 = \{ \psi_{\theta} : \ 0 \le \theta_1 \le \pi/4, \ 0 \le \theta_2 \le \pi/4 \}, \\ \mathcal{M}_6 = \{ \psi_{\theta} : \ 0 \le \theta_1 \le 2\pi/5, \ \theta_2 = 0 \}$$

Model \mathcal{M}_5 has the same amount of PQI as \mathcal{M}_4 , $J(\mathcal{M}_5) = \sqrt{3/6} = 0.289$. The PQI of the model \mathcal{M}_6 is $J(\mathcal{M}_6) = \frac{1}{2} \cos \frac{2\pi}{5} = 0.154$, which is smaller than $J(\mathcal{M}_5)$. This implies that \mathcal{M}_6 spreads more than \mathcal{M}_5 , and we see that the PQI is independent of the dimension of the parameter space.

Funding: This work was supported by JSPS KAKENHI Grant Numbers JP20H02168, JP19K11860.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Proofs

Appendix A.1. Proof of Lemma 7

For $\epsilon > 0$, we show that there exists *M* such that

$$|\theta| \ge M \Rightarrow |\langle f|g_{\theta} \rangle| < 2\sqrt{\epsilon}. \tag{A1}$$

First, we take compact sets *K* and *L* satisfying $\int_{K} |f|^{2} dx > 1 - \epsilon$ and $\int_{L} |g|^{2} dx > 1 - \epsilon$, respectively. Second, there exists a positive constant *M* such that $|\theta| \ge M \Rightarrow L + \theta = \{x + \theta : x \in L\}$ and *K* are disjoint. Note that $\int_{L+\theta} |g_{\theta}|^{2} dx > 1 - \epsilon$ for every θ . Then, we bound the absolute value of the inner product by two terms:

$$|\langle f|g_{\theta}\rangle| \leq \left|\int_{K} \bar{f}(x)g_{\theta}(x)dx\right| + \left|\int_{\sim K} \bar{f}(x)g_{\theta}(x)dx\right|,$$

where $\sim K$ denotes the complement of *K*. Due to the Cauchy–Schwarz inequality, the second term is bounded by

$$\begin{split} \left| \int_{\sim K} \bar{f}(x) g_{\theta}(x) dx \right| &\leq \sqrt{\int_{\sim K} |f|^2 dx} \sqrt{\int_{\sim K} |g_{\theta}|^2 dx} \\ &\leq \sqrt{\epsilon} \|g\|_2 \\ &= \sqrt{\epsilon}. \end{split}$$

The first term requires more steps. Since $K \subset \sim (L + \theta)$,

$$\begin{split} \left| \int_{K} \bar{f}(x) g_{\theta}(x) dx \right| &\leq \int_{K} |\bar{f}(x) g_{\theta}(x)| dx \\ &\leq \int_{\sim (L+\theta)} |\bar{f}(x) g_{\theta}(x)| dx \\ &\leq \sqrt{\int_{\sim (L+\theta)} |f|^{2} dx} \sqrt{\int_{\sim (L+\theta)} |g_{\theta}|^{2} dx} \\ &\leq \|f\|_{2} \sqrt{\epsilon} \\ &= \sqrt{\epsilon}. \end{split}$$

Putting them together, we obtain (A1). Therefore $|\langle g_{\theta} | f \rangle| \to 0$ as $|\theta| \to 0$ and we obtain (21) and (22).

Appendix A.2. Proof of Lemma 8

First from (A1), $|\langle g_{\theta}|g_{\theta'}\rangle| = |\langle g|g_{\theta'-\theta}\rangle| \to 0$ when $|\theta - \theta'| \to 0$. Thus, for $\epsilon > 0$, there exists *M* such that $|\theta - \theta'| \ge M \Rightarrow |\langle g_{\theta}|g_{\theta'}\rangle| < \epsilon$.

Using the above, we construct the sequence of a prior distribution with finite support (i.e., discrete probability) and Bayes mixture. First, for fixed *n*, we take a parameter set $\{\theta_1, \ldots, \theta_n\}$ satisfying $|\theta_i - \theta_j| \ge M$ for $i, j = 1, \ldots, n$.

Without loss of generality, we assume that $g_{\theta_1}, \ldots, g_{\theta_n}$ are linearly independent. Then, set $G_{ij} = \langle g_{\theta_i} | g_{\theta_j} \rangle$ ($G_{ii} = 1$). It is easily shown that the gram matrix $G = [G_{ij}]$ is positive definite. Then, we decompose *G* into the identity I_n and $A = G - I_n$, where diagonal components of *A* are zero.

For a uniform weight over the parameter set $\{\theta_1, \ldots, \theta_n\}$, we define the Bayes mixture

$$ho_n = rac{1}{n} \sum_{j=1}^n |g_{ heta_j}
angle \langle g_{ heta_j}|.$$

Now we show that

$$\|\rho_n\|_{\infty} = \frac{1}{n} \|G\|_{\infty} = \frac{1}{n} + \frac{1}{n} \|A\|_{\infty},$$
(A2)

$$\|A\|_{\infty} \le (n-1)\epsilon. \tag{A3}$$

First, we expand an arbitrary normalized vector ψ as $\psi = \sum_j c_j |g_{\theta_j}\rangle$. Then $1 = \langle \psi | \psi \rangle = c^{\dagger}Gc$, where $c = (c_1, \dots, c_n)^{\top}$ is a column vector and c^{\dagger} denotes the conjugate transpose of the vector *c*. Since *G* is positive definite, we take another parameter vector *d* as $d = \sqrt{Gc}$. Note that there is one-to-one correspondence between ψ and *d*. Thus,

$$egin{aligned} &\langle\psi|
ho_n|\psi
angle &=rac{1}{n}\sum_j|\langle\psi|g_{ heta_j}
angle|^2\ &=rac{1}{n}\|Gc\|_2^2\ &=rac{1}{n}d^{\dagger}Gd, \end{aligned}$$

where $||a||_2^2 = aa^{\dagger}$. This implies that

$$\begin{aligned} \|\rho_n\|_{\infty} &= \sup_{\psi: \|\psi\|_2 = 1} \langle \psi | \rho_n | \psi \rangle \\ &= \frac{1}{n} \sup_{d: \|d\|_2 = 1} d^{\dagger} G d \\ &= \frac{1}{n} + \frac{1}{n} \sup_{d: \|d\|_2 = 1} d^{\dagger} A d \\ &= \frac{1}{n} + \frac{1}{n} \|A\|_{\infty}, \end{aligned}$$

which shows Equation (A2).

Next, we show the inequality (A3). Due to Geršgorin's Theorem (see, e.g., Section 6.1 in Horn and Johnson [51]), all eigenvalues of A are located in the union of n discs

$$\bigcup_{i=1}^n \left\{ z \in \mathbf{C} : |z| \le \sum_{j \ne i} |G_{ij}| \right\}.$$

Thus, we easily show that the absolute value of each eigenvalue is bounded by $(n-1)\epsilon$.

Finally, we obtain from (A2) and (A3), $\|\rho_n\|_{\infty} \leq \frac{1}{n} + \epsilon$. For fixed $\epsilon > 0$, we have $0 \leq \inf_{\pi} \|\rho_{\pi}\|_{\infty} \leq \epsilon$. Since ϵ is arbitrary, $\inf_{\pi} \|\rho_{\pi}\|_{\infty}$ must be zero.

References

- 1. Iyer, P.; Poulin, D. A small quantum computer is needed to optimize fault-tolerant protocols. *Quantum Sci. Technol.* **2018**, *3*, 030504. [CrossRef]
- Kueng, R.; Long, D.M.; Doherty, A.C.; Flammia, S.T. Comparing experiments to the fault-tolerance threshold. *Phys. Rev. Lett.* 2016, 117, 170502. [CrossRef] [PubMed]
- 3. Holevo, A.S. Asymptotic Estimation of a Shift Parameter of a Quantum State. Theory Probab. Appl. 2005, 49, 207–220. [CrossRef]
- Fujiwara, A. Strong consistency and asymptotic efficiency for adaptive quantum estimation problems. J. Phys. A Math. Gen. 2006, 39, 12489–12504. [CrossRef]
- 5. Gill, R.D.; Massar, S. State estimation for large ensembles. *Phys. Rev. A* 2000, *61*, 042312. [CrossRef]
- 6. Hayashi, M. Asymptotic Theory of Quantum Statistical Inference; World Scientific: Singapore, 2005.
- 7. Helstrom, C.W. Quantum Detection Theory; Academic Press: New York, NY, USA, 1976.
- 8. Holevo, A.S. Probabilistic and Statistical Aspects of Quantum Theory; North-Holland: Amsterdam, The Netherlands, 1982.
- 9. Yamagata, K. Efficiency of quantum state tomography for qubits. Int. J. Quantum Inf. 2011, 9, 1167–1183. [CrossRef]
- 10. Eldar, Y.C. A semidefinite programming approach to optimal unambiguous discrimination of quantum states. *IEEE Trans. Inf. Theory* **2003**, *49*, 446–456. [CrossRef]
- 11. Eldar, Y.C.; Megretski, A.; Verghese, G.C. Designing optimal quantum detectors via semidefinite programming. *IEEE Trans. Inf. Theory* **2003**, *49*, 1007–1012. [CrossRef]
- 12. Eldar, Y.C.; Stojnic, M.; Hassibi, B. Optimal quantum detectors for unambiguous detection of mixed states. *Phys. Rev. A* 2004, *69*, 062318. [CrossRef]
- 13. Yuen, H.; Kennedy, R.; Lax, M. Optimal testing of multiple hypothesis in quantum detection theory. *IEEE Trans. Inf. Theory* **1975**, 2, 125–134. [CrossRef]
- 14. Blume-Kohout, R. Optimal, reliable estimation of quantum states. New J. Phys. 2010 12, 043034. [CrossRef]
- 15. Bužek, V.; Derka, R.; Adam, G.; Knight, P.L. Reconstruction of quantum states of spin systems: From quantum Bayesian inference to quantum tomography. *Ann. Phys.* **1998**, *266*, 454–496. [CrossRef]

- Smithey, D.T.; Beck, M.; Raymer, M.G.; Faridani, A. Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum. *Phys. Rev. Lett.* 1993, 70, 1244–1247. [CrossRef]
- 17. Vogel, K.; Risken, H. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A* **1989**, *40*, 2847–2849. [CrossRef]
- 18. Drezner, Z.; Wesolowsky, G.O. Facility location on a sphere. J. Opl. Res. Soc. 1978, 29, 997–1004. [CrossRef]
- 19. Shiode, S. Minimax facility location problem on a sphere. Rev. Kobe Univ. Mercantile Mar. 1989, 37, 155–158.
- 20. Tanaka, F. Noninformative prior in the quantum statistical model of pure states. Phys. Rev. A 2012, 85, 062305. [CrossRef]
- 21. van Dam, W.; Hayden, P. Rényi-entropic bounds on quantum communication. arXiv 2002, arXiv:quant-ph/0204093.
- 22. Hu, X.; Ye, Z. Generalized quantum entropy. J. Math. Phys. 2006, 47, 023502. [CrossRef]
- 23. Davisson, L.; Leon-Garcia, A. A source matching approach to finding minimax codes. *IEEE Trans. Inf. Theory* **1980**, *26*, 166–174. [CrossRef]
- 24. Brody, D.C.; Hook, D.W. On optimum Hamiltonians for state transformation. J. Phys. A Math. Theor. 2007, 40, 10949. [CrossRef]
- 25. Cafaro, C.; Ray, S.; Alsing, P.M. Geometric aspects of analog quantum search evolutions. Phys. Rev. A 2020, 102, 052607. [CrossRef]
- Laba, H.P.; Tkachuk, V.M. Geometric characteristics of quantum evolution: Curvature and torsion. *Condensed Matter Phys.* 2017, 20, 13003. [CrossRef]
- 27. Mostafazadeh, A. Hamiltonians generating optimal-speed evolutions. Phys. Rev. A 2009, 79, 014101. [CrossRef]
- 28. Tanaka, F. Bayesian estimation of the wave function. Phys. Lett. A 2012, 376, 2471–2476. [CrossRef]
- 29. Bengtsson, I.; Życzkowki, K. Geometry of Quantum States; Cambridge University Press: Cambridge, UK, 2006.
- 30. Ferguson, T. Mathematical Statistics: A Decision Theoretic Approach; Academic Press: New York, NY, USA, 1967.
- 31. Berger, J. Statistical Decision Theory and Bayesian Analysis, 2nd ed.; Springer: New York, NY, USA, 1985.
- 32. Bernardo, J.M.; Smith, A.F.M. Bayesian Theory; Wiley: Hoboken, NJ, USA, 1994.
- 33. Robert, C.P. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation;* Springer: New York, NY, USA, 2001.
- Arimoto, S. An algorithm for computing the capacity of arbitrary memoryless channels. *IEEE Trans. Inf. Theory* 1972, 18, 14–20. [CrossRef]
- 35. Blahut, R.E. Computation of channel capacity and rate distortion functions. IEEE Trans. Inf. Theory 1972, 18, 460–473. [CrossRef]
- 36. Nielsen, M.A.; Chuang, I.L. Quantum Computation and Quantum Information; Cambridge University Press: Cambridge, UK, 2000.
- 37. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley-Interscience: New York, NY, USA, 2005.
- Rényi, A. On measures of entropy and information. In Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, California, CA, USA, 20 June–30 July 1960; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547–561.
- 39. Shannon, C.E. A mathematical theory of communication. Bell Syst. Tech. J. 1948, 27, 379–423, 623–656. [CrossRef]
- 40. Tsallis, C. Introduction to Nonextensive Statistical Mechanics; Springer: New York, NY, USA, 2010.
- 41. Ohya, M.; Petz, D. Quantum Entropy and Its Use; Springer: Berlin/Heidelberg, Germany, 2004.
- 42. Jaynes, E.T. Information theory and statistical mechanics. I. Phys. Rev. 1957, 106, 620–630. [CrossRef]
- 43. Jaynes, E.T. Information theory and statistical mechanics. II. Phys. Rev. 1957, 108, 171–190. [CrossRef]
- 44. Kapur, J.N.; Kesavan, H.K. Entropy Optimization Principles with Applications; Academic Press: Boston, MA, USA, 1992.
- 45. Rubinstein, R.Y.; Kroese, D.P. The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning; Springer: New York, NY, USA, 2004.
- 46. Ruskai, M.B. Extremal properties of relative entropy in quantum statistical mechanics. *Rep. Math. Phys.* **1988**, *26*, 143–150. [CrossRef]
- 47. Brown, L.D. Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory; Institute of Mathematical Statistics, Hayward: Berkeley Heights, NJ, USA, 1986.
- 48. Barndorff-Nielsen, O.E. Information and Exponential Families in Statistical Theory; Wiley: New York, NY, USA, 1978.
- Goyal, P.; Giffin, A.; Knuth, K.H.; Vrscay, E. (Eds.) Bayesian Inference and Maximum Entropy Methods in Science and Engineering; AIP Conference Series; AIP Publishing: Melville, NY, USA, 2012; p. 1443.
- 50. Lehmann, E.L. Theory of Point Estimation; Wiley: New York, NY, USA, 1983.
- 51. Horn, R.A.; Johnson, C.R. Matrix Analysis; Cambridge University Press: New York, NY, USA, 1985.