2025/4/7

Heisenberg scaling based on population coding

Masahito Hayashi

1 State estimation

To see the standard scaling for the mutual information, we consider state estimation, i.e., we focus on a d-parameter state family $\{\rho_\theta\}_{\theta\in\Theta}$ on the Hilbert space $\mathcal{H}_A$, where Θ is a subset of $\mathbb{R}^d$. We consider a Bayesian prior μ on Θ. We denote the set of density matrices on $\mathcal{H}$ by $\mathcal{S}(\mathcal{H})$. We denote the classical system Θ by B. We consider n-copy state $\rho_\theta^{\otimes n}$. We have classical-quantum state

$$\rho_{AB} := \int_\Theta \rho_\theta^{\otimes n} \otimes |\theta\rangle\langle\theta| \mu(d\theta). \tag{1.1}$$

We focus on the mutual information between A and B, which is given as

$$I(A;B) = S(\rho_A) - \int_\Theta S(\rho_\theta^{\otimes n})\mu(d\theta) = \int_\Theta D(\rho_\theta^{\otimes n}\|\rho_A)\mu(d\theta), \tag{1.2}$$

where $S(\rho)$ expresses the von Neumann entropy $-\mathrm{Tr}\,\rho\log\rho$ of ρ, and $D(\rho\|\sigma)$ expresses the quantum relative entropy $\mathrm{Tr}\,\rho(\log\rho - \log\sigma)$.
When all densities are commutative, i.e., the model is classical, the references [31, 32] showed that

$$I(A;B) = \frac{d}{2}\log n + O(1). \tag{1.3}$$

When our model $\{\rho_\theta\}_{\theta\in\Theta}$ is the full model on t-dimensional system and μ is invariant for unitary action, the reference [33] showed that

$$I(A;B) = \frac{t^2-1}{2}\log n + O(1). \tag{1.4}$$

Since the number of parameters of the full model is t^2-1, (1.4) can be considered as a generalization of (1.3). When a model satisfies a certain condition, using the result by [36], we can show

$$I(A;B) = \frac{d}{2}\log n + o(\log n). \tag{1.5}$$

That is, the leading term is the number of parameters times $\frac{1}{2}\log n$.

We back to the spirit of population coding, and focus on the number $N(\{\rho_\theta^{\otimes n}\}_{\theta\in\Theta},\epsilon)$ of distinguishable states among $\{\rho_\theta^{\otimes n}\}_{\theta\in\Theta}$ with the average decoding error probability $\epsilon > 0$. As shown in [44, (4.32)], using Fano inequality, we can evaluate this number as

$$\log N(\{\rho_\theta^{\otimes n}\}_{\theta\in\Theta},\epsilon) \le \frac{\log 2 + I(A;B)}{1-\epsilon}. \tag{1.6}$$

That is, the relations (1.3), (1.4), and (1.5) give upper bounds of this number.

2 General problem formulation

2.1 General description

Let G be a compact group and μ be the Haar measure of G. We focus on a unitary representation f of G over a finite-dimensional Hilbert space. For this aim, we choose an input state. We assume that any entangled input state with any reference system is available. To consider this problem, we denote the set of the labels of irreducible representations of G by $\hat{G}$. Let $\mathcal{U}_\lambda$ be the irreducible representation space identified by $\lambda \in \hat{G}$, and d_λ be its dimension.

We also define the twirling operation for the group G as

$$\mathcal{T}_G(\rho) := \int_G f(g')\rho f(g')^\dagger \mu(dg'). \tag{2.1}$$

We denote the set of irreducible representations appearing in f by $\hat{G}_f$. We decompose the representation system $\mathcal{H}_A$ as follows.

$$\mathcal{H}_A = \bigoplus_{\lambda \in \hat{G}_f} \mathcal{U}_\lambda \otimes \mathbb{C}^{n_\lambda}, \tag{2.2}$$

where n_λ expresses the multiplicity of the representation space $\mathcal{U}_\lambda$.

When a reference system $\mathbb{C}^l$ is available, we have

$$\mathcal{H}_{Al} := \mathcal{H}_A \otimes \mathbb{C}^l = \bigoplus_{\lambda \in \hat{G}_f} \mathcal{U}_\lambda \otimes \mathbb{C}^{l n_\lambda}. \tag{2.3}$$

However, when the input state is a pure state, the orbit is restricted to the following space by choosing a suitable subspace $\mathbb{C}^{\min(d_\lambda, l n_\lambda)}$ of $\mathbb{C}^{l n_\lambda}$. That is, our representation space can be considered as follows.

$$\bigoplus_{\lambda \in \hat{G}_f} \mathcal{U}_\lambda \otimes \mathbb{C}^{\min(d_\lambda, l n_\lambda)}. \tag{2.4}$$

When $l \geq d_\lambda / n_\lambda$ for any $\lambda \in \hat{G}_f$, our representation is given as

$$\mathcal{H}_{AR} := \bigoplus_{\lambda \in \hat{G}_f} \mathcal{U}_\lambda \otimes \mathbb{C}^{d_\lambda}. \tag{2.5}$$

In the following, we consider the above case. We denote the projection to $\mathcal{U}_\lambda \otimes \mathbb{C}^{d_\lambda}$ by P_λ.

2.2 Mutual information

We choose an input state ρ and a distribution P on the system $\mathcal{B} = G$, where the random choice of $\mathcal{B}$ is denoted by the random variable B. So, we have a classical-quantum state $\sum_{g\in\mathcal{X}} P(g) f(g)\rho f(g)^\dagger \otimes |g\rangle\langle g|$. We focus on the mutual information $I(Al;B)$ or $(AR;B)$, which depends on ρ and P. When ρ is decomposed as $\sum_j q_j \rho_j$, the mutual information is evaluated as

$$\begin{aligned} I(Al;B)[P,\rho] :=& S\Big(\sum_{g'\in\mathcal{X}} P(g')f(g')\rho f(g')^\dagger\Big) - \sum_{g\in\mathcal{X}} P(g) S\Big(f(g)\rho f(g)^\dagger\Big) \\ =& \sum_{g\in\mathcal{X}} P(g) D\Big(f(g)\rho f(g)^\dagger \Big\| \sum_{g'\in\mathcal{X}} P(g')f(g')\rho f(g')^\dagger\Big) \\ \le& \sum_j q_j \sum_{g\in\mathcal{X}} P(g) D\Big(f(g)\rho_j f(g)^\dagger \Big\| \sum_{g'\in\mathcal{X}} P(g')f(g')\rho_j f(g')^\dagger\Big). \end{aligned} \tag{2.6}$$

The final inequality follows from the joint convexity of relative entropy. Since any mixed state can be written as a mixture of pure states, to maximize the mutual information, we can restrict the input state as an input pure state $|\psi\rangle = \oplus_{\lambda\in\hat{G}_f} \sqrt{p_\lambda}|\psi_\lambda\rangle$. In this case, since the von Neumann entropy of a pure state is zero, the mutual information is given as the von Neumann entropy of the mixture of the final states.

$$I(Al;B)[P,|\psi\rangle] = S\Big(\sum_{g'\in\mathcal{X}} P(g')f(g')|\psi\rangle\langle\psi|f(g')^\dagger\Big). \tag{2.7}$$

We consider the following optimization problem.

$$I(G)_{Al} := \max_{|\psi\rangle}\max_{P} I(Al;B)[P,|\psi\rangle]. \tag{2.8}$$

As shown in [45], due to the convexity of von Neumann entropy the maximum is attained by the Haar measure μ. Hence,

$$\max_{P} I(Al;B)[P,|\psi\rangle] = I(Al;B)[\mu,|\psi\rangle] = S(\mathcal{T}_G(|\psi\rangle\langle\psi|)), \tag{2.9}$$

which implies

$$I(G)_{Al} = \max_{|\psi\rangle} S(\mathcal{T}_G(|\psi\rangle\langle\psi|)). \tag{2.10}$$

Here, we set $|\phi_\lambda\rangle$ to be a pure state $\mathcal{U}_\lambda \otimes \mathbb{C}^{\min(d_\lambda, ln_\lambda)}$ such that $\mathrm{Tr}_{\mathcal{U}_\lambda} |\phi_\lambda\rangle\langle\phi_\lambda|$ is the completely mixed state on $\mathbb{C}^{\min(d_\lambda, ln_\lambda)}$. Such a state $|\phi_\lambda\rangle$ is called the maximally entangled state on $\mathcal{U}_\lambda \otimes \mathbb{C}^{\min(d_\lambda, ln_\lambda)}$.

When $|\psi\rangle$ is the state $|\phi(p)\rangle := \sum_{\lambda \in \hat{G}_f} \sqrt{p_\lambda}|\phi_\lambda\rangle$, where p is a distribution on $\hat{G}_f$, we have

$$S(\mathcal{T}_G(|\phi(p)\rangle\langle\phi(p)|)) = S\Big(\sum_{\lambda \in \hat{G}_f} p_\lambda \frac{P_\lambda}{d_\lambda \min(n_\lambda, d_\lambda)}\Big)$$
$$= S(p) + \sum_{\lambda \in \hat{G}_f} p_\lambda \log(d_\lambda \min(n_\lambda, d_\lambda)). \tag{2.11}$$

Since the dimension of support of $\mathcal{T}_G(|\psi\rangle\langle\psi|)$ is upper bounded by $\sum_{\lambda \in \hat{G}} d_\lambda \min(n_\lambda, d_\lambda)$, we have

$$I(X; Al)[\mu, |\psi\rangle] = S(\mathcal{T}_G(|\psi\rangle\langle\psi|)) \leq R(G)_{Al} := \log\Big(\sum_{\lambda \in \hat{G}_f} d_\lambda \min(ln_\lambda, d_\lambda)\Big). \tag{2.12}$$

The equality holds when the input state is $|\phi(p)\rangle$ and $p_\lambda = \frac{d_\lambda \min(ln_\lambda, d_\lambda)}{\sum_{\lambda' \in \hat{G}_f} d_{\lambda'} \min(ln_{\lambda'}, d_{\lambda'})}$. Therefore, we have

$$I(G)_{Al} = \log\Big(\sum_{\lambda \in \hat{G}_f} d_\lambda \min(ln_\lambda, d_\lambda)\Big). \tag{2.13}$$

Overall, the optimal input state is

$$\sum_{\lambda \in \hat{G}_f} \sqrt{\frac{d_\lambda \min(ln_\lambda, d_\lambda)}{\sum_{\lambda' \in \hat{G}_f} d_{\lambda'} \min(ln_{\lambda'}, d_{\lambda'})}} |\phi_\lambda\rangle. \tag{2.14}$$

In particular, when l satisfies the condition $l \geq d_\lambda / n_\lambda$ for any $\lambda \in \hat{G}_f$,

$$I(G)_{AR} = I(G)_{Al} = R(G)_{AR} := \log\Big(\sum_{\lambda \in \hat{G}_f} d_\lambda^2\Big). \tag{2.15}$$

2.3 Number of distinguishable elements

Next, we consider how many elements in G can be distinguished via the above process. When the input state is ρ, we denote the number of distinguishable states among $\{f(g)\rho f(g)^\dagger\}_{g\in G}$ with the average decoding error $\epsilon > 0$ by $M_{Al}(\rho,\epsilon)$. That is,

$$M_{Al}(\rho,\epsilon) := \max\left\{ M \middle| \begin{array}{l} \frac{1}{M}\sum_{j=1}^{M} \mathrm{Tr}\, f(g_j)\rho f(g_j)^\dagger \Pi_j \\ \geq 1-\epsilon \end{array} \right\}, \tag{2.16}$$

where the maximum in (2.16) is taken with respect to $M, g_1, \ldots, g_M \in G, \{\Pi_j\}_{j=1}^M$. We consider the following optimization problems.

$$M_\epsilon(G)_{Al} := \max_\rho M_{Al}(\rho,\epsilon). \tag{2.17}$$

We have

$$\log M(|\psi\rangle,\epsilon) \geq \max_P S_\alpha\Big(\sum_{g'\in\mathcal{X}} P(g')f(g')|\psi\rangle\langle\psi|f(g')^\dagger\Big) - \frac{1}{\alpha-1}(\log 2 - \log\epsilon), \tag{2.18}$$

$$\log M(|\psi\rangle,\epsilon) \leq \max_P S_\beta\Big(\sum_{g'\in\mathcal{X}} P(g')f(g')|\psi\rangle\langle\psi|f(g')^\dagger\Big) + \frac{1}{\beta-1}\log(1-\epsilon) \tag{2.19}$$

for $0 < \beta < 1$ and $1 < \alpha \leq 2$. For its detailed derivation, see Appendix AppendixA. The above maximum is attained when P is the Haar measure μ.
Since the dimension of support of $\mathcal{T}_G(|\psi\rangle\langle\psi|)$ is upper bounded by $\sum_{\lambda \in \hat{G}} d_\lambda \min(n_\lambda, d_\lambda)$, we have

$$S_\alpha(\mathcal{T}_G(|\psi\rangle\langle\psi|)) \leq R(G)_{Al}. \tag{2.20}$$

The equality holds when the input state is $|\phi\rangle$. Thus, choosing the best choice $\alpha = 2$ and taking the limit $\beta \to +0$, we have

$$R(G)_{Al} - (\log 2 - \log \epsilon) \leq \log M_\epsilon(G)_{Al} \leq R(G)_{Al} - \log(1-\epsilon). \tag{2.21}$$

In particular, we have

$$R(G)_{AR} - (\log 2 - \log \epsilon) \leq \log M_\epsilon(G)_{AR} \leq \log R(G)_{AR} - \log(1-\epsilon). \tag{2.22}$$

Therefore, when $1 > \epsilon > 0$ is fixed, $\log R(G)_{AR}$ is the dominant term.

3 Multiple phase estimation

Here, we consider the multi-phase application model on the t-dimensional system $\mathcal{H}_A$ spanned by $\{|j\rangle\}_{j=0}^{t-1}$. This model is given as the application of the group $MP_{t-1} := \{U_\theta\}_{\theta\in[0,2\pi)^{t-1}}$, where the unitary U_θ is defined as $|0\rangle\langle 0| + \sum_{j=1}^{t-1} e^{i\theta_j}|j\rangle\langle j|$. Now, we consider the n-parallel application of U_θ on $\mathcal{H}_{A^n} := \mathcal{H}_A^{\otimes n}$. In this representation, all irreducible representations are characterized by the combinatorics $\mathcal{C}_t^n$, which is defined as

$$\mathcal{C}_t^n := \left\{ \vec{n} := (n_0, n_1, \dots, n_{t-1}) \in \mathbb{N}^t \middle| \sum_{j=0}^{t-1} n_j = n \right\}, \tag{3.1}$$

where $\mathbb{N}$ is the set of non-negative integers. Since all irreducible representations are one-dimensional, $d_{\vec{n}} = 1$ for $\vec{n} \in \mathcal{C}_t^n$ so that we do not need to attach any reference system, i.e., $R(G)_{A1} = R(G)_{AR}$. Since $|\mathcal{C}_t^n| = \binom{n+t-1}{t-1}$, we have

$$I(MP_{t-1})_{A^n} = R(MP_{t-1})_{A^n} = \log \binom{n+t-1}{t-1} = (t-1)\log n + o(1) \tag{3.2}$$

while the number of our parameters is $t-1$. Compare with (1.5), (3.2) achieves the twice of the state estimation case.

When $t=2$, the optimal input state is

$$\sum_{j=0}^{n} \frac{1}{\sqrt{n+1}} |\overbrace{0\dots0}^{j}\overbrace{1\dots1}^{n-j}\rangle. \tag{3.3}$$

The asymptotic estimation error with the above input state has been studied in [46, Sections 3 and 4]. As explained in [19, Section 4], this method is essentially the same as Kitaev's method [48, 49], and has been implemented by [50].
Here, we compare the input state (3.3) with the case with the noon state [9, 8, 7].
The input noon state has only two irreducible representations; the subspace spanned by $|0\rangle^{\otimes n}$ and $|1\rangle^{\otimes n}$. The relations (2.9) and (2.11) imply

$$I(X;A1)[\mu, \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})] = \log 2. \tag{3.4}$$

The input noon state has quite small global information, which certifies the impossibility of global estimation by the input noon state.
When we focus on the error $\sin^2 \frac{\theta_{guess}-\theta_{true}}{2}$, the input state (3.3) does not coincide with the optimal input. Further, the input state (3.3) does not achieve Heisenberg limit in this sense, but the input state (3.3) achieves the Heisenberg scaling in the sense limiting distribution.

4 n-tensor product representation of $\mathrm{SU}(t)$ on $\mathcal{H}_A = \mathbb{C}^t$

Y_t^n: Set of Semistandard Young tableau.
$\mathcal{M}_\lambda = \mathbb{C}^{n_\lambda}$: Multiplicity space of Irreducible Unitary Representation U^λ with Multiplicity n_λ.
Our representation space:

$$\mathcal{H}_{A^n} = \bigoplus_{\lambda \in Y_t^n} \mathcal{H}_\lambda \otimes \mathcal{M}_\lambda. \tag{4.1}$$

When $\lambda = (\lambda_1, \ldots, \lambda_t)$ with $\lambda_1 \le \ldots \le \lambda_t$ and $\sum_j \lambda_j = n$, $d_\lambda := \dim \mathcal{H}_\lambda$ is given as

$$d_\lambda = \prod_{1 \le j < k \le t} \frac{k - j + \lambda_k - \lambda_j}{k - j}. \tag{4.2}$$

It is upper bounded by $(n+1)^{t(t-1)/2}$. When the reference system $\mathbb{C}^l$ is sufficiently large in the sense of (2.5), the maximum mutual information is evaluated as

$$\begin{aligned} I(\mathrm{SU}(t))_{A^n R} = R(\mathrm{SU}(t))_{A^n R} = \log \Big(\sum_{\lambda' \in Y_t^n} d_{\lambda'}^2 \Big) \\ \le \log (n+1)^{(t+1)t + (t-1)} = (t^2 - 1) \log(n+1). \end{aligned} \tag{4.3}$$

Due to (2.14), the optimal input state is

$$\sum_{\lambda \in Y_t^n} d_\lambda \sqrt{\Big(\sum_{\lambda' \in Y_t^n} d_{\lambda'}^2\Big)^{-1}} |\phi_\lambda\rangle, \tag{4.4}$$

Here, $|\phi_\lambda\rangle$ is the maximally entangled state on $\mathcal{H}_\lambda \otimes \mathcal{M}_\lambda$.
When t is fixed and only n increases, we have

$$\log\Big(\sum_{\lambda' \in Y_t^n} d_{\lambda'}^2\Big) = (t^2 - 1)\log n + O(1), \tag{4.5}$$

which is shown in Section 5. Since $t^2 - 1$ is the number of parameters of $\mathrm{SU}(t)$. The leading term in (4.5) is twice of the case of state estimation case given in (1.5). Therefore, this behavior can be considered to achieve the Heisenberg scaling in the sense of mutual information.

At least, when $t = 2$, we can show the above relation as follows. Consider the case with $t = 2$. When $n = 2m$, we have

$$\sum_{\lambda' \in Y_d^n} d_{\lambda'}^2 = \sum_{k=0}^{m} (2k+1)^2 = \frac{(m+1)(2m+1)(2m+3)}{3}. \tag{4.6}$$

When $n = 2m - 1$, we have

$$\sum_{\lambda' \in Y_d^n} d_{\lambda'}^2 = \sum_{k=1}^{m} (2k)^2 = \frac{2m(m+1)(2m+1)}{3}. \tag{4.7}$$

Then, we have

$$\log\Big(\sum_{\lambda' \in Y_d^n} d_{\lambda'}^2\Big) = 3\log n - \log 6 + O(\frac{1}{n}). \tag{4.8}$$

Here, we compare the above optimal case (4.3),(4.5) with the case of the input state maximizing the trace of Fisher information matrix [11]. The maximally entangled state $|\Phi\rangle$ on the symmetric subspace maximizes the trace of Fisher information matrix [11]. In this case, since the dimension of the symmetric subspace is $\binom{n+t-1}{t-1}$, the relations (2.9) and (2.11) imply

$$I(X;AR)[\mu, |\Phi\rangle] = 2\log\binom{n+t-1}{t-1} = 2(t-1)\log n + o(1), \tag{4.9}$$

which is much smaller than (4.3),(4.5).
Next, we focus on $M_\epsilon(\mathrm{SU}(t))_{AR}$. Due to (2.22), (4.3), and (4.5), $\log M_\epsilon(\mathrm{SU}(t))_{AR}$ behaves as

$$\log M_\epsilon(\mathrm{SU}(t))_{AR} = (t^2-1)\log n + O(1). \tag{4.10}$$

Since $\mathrm{SU}(t)$ is parametrized by t^2-1 parameters, $B(R)$ scales as $O(R^{t^2-1})$. Since $M_\epsilon(\mathrm{SU}(t))_{AR}$ scales as $O(n^{t^2-1})$, the upper bound of R_ϵ scales as $O(n^{-1})$.

5 Proof of (4.5)

Since we have (4.3), it is sufficient to show

$$\log\Big(\sum_{\lambda'\in Y_t^n} d_{\lambda'}^2\Big) \geq (t^2-1)\log n + O(1). \tag{5.1}$$

We choose a positive real number a_n such that $\frac{n}{a_n t}$ is an integer. We define the subset $Y_t^n(a_n) \subset Y_t^n$ as

$$Y_t^n(a_n) := \Big\{\lambda \Big| \frac{n}{a_n t} \leq \lambda_{k+1} - \lambda_k \text{ for } k = 1, \dots, t-1\Big\}. \tag{5.2}$$

For $\lambda \in Y_t^n(a_n)$, we have

$$\begin{aligned} d_\lambda &= \prod_{1\leq j<k\leq t} \frac{k-j+\lambda_k-\lambda_j}{k-j} \\ &\geq \frac{1}{t} \prod_{1\leq j<k\leq t} \frac{\lambda_k-\lambda_j}{t} \geq \prod_{1\leq j<k\leq t} \frac{n}{a_n t^2} = \big(\frac{n}{a_n t^2}\big)^{\frac{(t-1)t}{2}}. \end{aligned} \tag{5.3}$$

Also, using $m_n := n - \frac{n}{a_n}$, we have

$$|Y_t^n(a_n)| = \binom{m_n+t-1}{t-1} \geq \frac{m_n^{t-1}}{(t-1)}. \tag{5.4}$$

Therefore,

$$\sum_{\lambda \in Y_t^n} d_\lambda^2 \geq \sum_{\lambda \in Y_t^n(a_n)} d_\lambda^2 \geq \sum_{\lambda \in Y_t^n(a_n)} \big(\frac{n}{a_n t^2}\big)^{(t-1)t}$$
$$= |Y_t^n(a_n)| \big(\frac{n}{a_n t^2}\big)^{(t-1)t} \geq \frac{m_n^{t-1}}{(t-1)} \big(\frac{n}{a_n t^2}\big)^{(t-1)t}. \tag{5.5}$$

When a_n choose as a value such that $2 \leq a_n \leq 3$, we have

$$\log\left(\frac{m_n^{t-1}}{(t-1)} \big(\frac{n}{a_n t^2}\big)^{(t-1)t}\right) = (t^2-1)\log n + O(1) \tag{5.6}$$

because $t - 1 + (t-1)t = t^2 - 1$. Hence, we obtain (5.1).

6 Conclusion

We have derived general formulas for the mutual information and the logarithm of the number of distinguishable elements when a unitary group representation is given. We have proposed the above quantities as a figure of merit to address the population coding with group representation because these quantities reflect the global information structure unlike Fisher information. Then, we have applied these general formulas to the case with multi-phase estimation and multiple applications of $\mathrm{SU}(t)$. As the results, we have revealed that the optimal strategy realized the twice value of the standard case for these two quantities, which can be considered as Heisenberg scaling. We have also shown that the optimal strategies for maximizing Fisher information have much smaller values for these quantities. This fact shows the advantage of our figure of merit over Fisher information.

AppendixA Derivations of (2.18) and (2.19)

Since $f(g')|\psi\rangle\langle\psi|f(g')^\dagger$ is a pure state, the reference [47, (6)] shows the inequality

$$\begin{aligned}\epsilon \le & 2(M(|\psi\rangle,\epsilon)-1)^s \operatorname{Tr}\Big(\sum_{g'\in\mathcal{X}} P(g')f(g')|\psi\rangle\langle\psi|f(g')^\dagger\Big)^{1+s} \\ \le & 2M(|\psi\rangle,\epsilon)^s \operatorname{Tr}\Big(\sum_{g'\in\mathcal{X}} P(g')f(g')|\psi\rangle\langle\psi|f(g')^\dagger\Big)^{1+s} \end{aligned} \quad \text{(AppendixA.1)}$$

for $0 \le s \le 1$ and any distribution P on G. Choosing $\alpha = 1+s$, we have

$$\log M(|\psi\rangle,\epsilon) \ge S_\alpha\Big(\sum_{g'\in\mathcal{X}} P(g')f(g')|\psi\rangle\langle\psi|f(g')^\dagger\Big) - \frac{1}{\alpha-1}(\log 2 - \log\epsilon) \quad \text{(AppendixA.2)}$$

for $1 < \alpha \le 2$. Taking the maximum for P, we obtain (2.18).
Since $f(g')|\psi\rangle\langle\psi|f(g')^\dagger$ is a pure state, the reference [44, (4.67)] shows the inequality

$$\begin{aligned}&\log(1-\epsilon) \\ \le & \max_P \log \operatorname{Tr}\Big(\sum_{g'\in\mathcal{X}} P(g')f(g')|\psi\rangle\langle\psi|f(g')^\dagger\Big)^{\frac{1}{1-s}} + \frac{s}{1-s}\log M(|\psi\rangle,\epsilon)\end{aligned} \quad \text{(AppendixA.3)}$$

for $s < 0$. Choosing $\beta = \frac{1}{1-s}$, we have (2.19).

References

[1] Carl Wilhelm Helstrom. "Quantum detection and estimation theory". Academic press. (1976).

[2] Alexander S Holevo. "Probabilistic and statistical aspects of quantum theory". Edizioni della Normale. (2011).

[3] Richard D. Gill and Serge Massar. "State estimation for large ensembles". Phys. Rev. A **61**, 042312 (2000).

[4] Yuxiang Yang, Giulio Chiribella, and Masahito Hayashi. "Attaining the ultimate precision limit in quantum state estimation". Communications in Mathematical Physics **368**, 223–293 (2019).

[5] Masahito Hayashi and Keiji Matsumoto. "Asymptotic performance of optimal state estimation in qubit system". Journal of Mathematical Physics **49**, 102101 (2008).

[6] Masahito Hayashi and Yingkai Ouyang. "Tight Cramér-Rao type bounds for multiparameter quantum metrology through conic programming". Quantum **7**, 1094 (2023).

[7] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Advances in quantum metrology". Nature Photonics **5**, 222–229 (2011).

[8] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum metrology".

Phys. Rev. Lett. **96**, 010401 (2006).

[9] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum-enhanced measurements: Beating the standard quantum limit". Science **306**, 1330–1336 (2004).

[10] Jonathan A. Jones, Steven D. Karlen, Joseph Fitzsimons, Arzhang Ardavan, Simon C. Benjamin, G. Andrew D. Briggs, and John J. L. Morton. "Magnetic field sensing beyond the standard quantum limit using 10-spin noon states". Science **324**, 1166–1168 (2009).

[11] Hiroshi Imai and Akio Fujiwara. "Geometry of optimal estimation scheme for su(d) channels". Journal of Physics A: Mathematical and Theoretical **40**, 4391 (2007).

[12] Ryo Okamoto, Holger F Hofmann, Tomohisa Nagata, Jeremy L O'Brien, Keiji Sasaki, and Shigeki Takeuchi. "Beating the standard quantum limit: phase super-sensitivity of n-photon interferometers". New Journal of Physics **10**, 073033 (2008).

[13] Tomohisa Nagata, Ryo Okamoto, Jeremy L. O'Brien, Keiji Sasaki, and Shigeki Takeuchi. "Beating the standard quantum limit with four-entangled photons". Science **316**, 726–729 (2007).

[14] Nicholas Thomas-Peter, Brian J. Smith, Animesh Datta, Lijian Zhang, Uwe Dorner, and Ian A. Walmsley. "Real-world quantum sensors: Evaluating resources

for precision measurement". Phys. Rev. Lett. **107**, 113603 (2011).
[15] V. Bužek, R. Derka, and S. Massar. "Optimal quantum clocks". Phys. Rev. Lett. **82**, 2207–2210 (1999).
[16] Masahito Hayashi. "Parallel treatment of estimation of SU(2) and phase estimation". Physics Letters A **354**, 183–189 (2006).
[17] Masahito Hayashi. "Comparison between the cramer-rao and the mini-max approaches in quantum channel estimation". Communications in Mathematical Physics **304**, 689–709 (2011).
[18] Masahito Hayashi, Sai Vinjanampathy, and L C Kwek. "Resolving unattainable cramer-rao bounds for quantum sensors". Journal of Physics B: Atomic, Molecular and Optical Physics **52**, 015503 (2018).
[19] Masahito Hayashi, Zi-Wen Liu, and Haidong Yuan. "Global heisenberg scaling in noisy and practical phase estimation". Quantum Science and Technology **7**, 025030 (2022).
[20] Wojciech Górecki, Rafał Demkowicz-Dobrzański, Howard M Wiseman, and Dominic W Berry. "π-corrected heisenberg limit". Physical review letters **124**, 030501 (2020).
[21] Raphael Kaubruegger, Denis V. Vasilyev, Marius Schulte, Klemens Hammerer, and Peter Zoller. "Quantum variational optimization of ramsey interferometry and atomic clocks". Phys. Rev. X **11**, 041045 (2021).

[22] Xinglei Yu, Xinzhi Zhao, Liangsheng Li, Xiao-Min Hu, Xiangmei Duan, Haidong Yuan, and Chengjie Zhang. "Toward heisenberg scaling in non-hermitian metrology at the quantum regime". Science Advances **10**, eadk7616 (2024).
[23] Qiushi Liu, Zihao Hu, Haidong Yuan, and Yuxiang Yang. "Optimal strategies of quantum metrology with a strict hierarchy". Phys. Rev. Lett. **130**, 070803 (2023).
[24] Ran Liu, Yu Chen, Min Jiang, Xiaodong Yang, Ze Wu, Yuchen Li, Haidong Yuan, Xinhua Peng, and Jiangfeng Du. "Experimental critical quantum metrology with the heisenberg scaling". npj Quantum Information **7**, 170 (2021).
[25] Jingyi Fan and Shengshi Pang. "Achieving heisenberg scaling by probe-ancilla interaction in quantum metrology" (2024).
[26] Le Hu, Shengshi Pang, and Andrew N. Jordan. "Achieving heisenberg scaling on measurement of a three-qubit system via quantum error correction". Phys. Rev. A **106**, 052609 (2022).
[27] Lupei Qin, Jialin Li, Yazhi Niu, and Xin-Qi Li. "Enhanced super-heisenberg scaling precision by nonlinear coupling and postselection". Physics Letters A **523**, 129795 (2024).
[28] Danilo Triggiani, Paolo Facchi, and Vincenzo Tamma. "Heisenberg scaling precision in the estimation of functions of parameters in linear optical networks". Phys. Rev. A **104**, 062603 (2021).

[29] Si Wu, Shun-ichi Amari, and Hiroyuki Nakahara. "Population Coding and Decoding in a Neural Field: A Computational Study". Neural Computation **14**, 999–1026 (2002).

[30] Stefano Panzeri, Fernando Montani, Giuseppe Notaro, Cesare Magri, and Rasmus S. Peterson. "Population coding". Pages 303–319. Springer US. Boston, MA (2010).

[31] B.S. Clarke and A.R. Barron. "Information-theoretic asymptotics of bayes methods". IEEE Transactions on Information Theory **36**, 453–471 (1990).

[32] Bertrand S. Clarke and Andrew R. Barron. "Jeffreys' prior is asymptotically least favorable under entropy risk". Journal of Statistical Planning and Inference **41**, 37–60 (1994).

[33] Masahito Hayashi. "Universal approximation of multi-copy states and universal quantum lossless data compression". Communications in Mathematical Physics **293**, 171–183 (2010).

[34] Yuxiang Yang, Giulio Chiribella, and Masahito Hayashi. "Communication cost of quantum processes". IEEE Journal on Selected Areas in Information Theory **1**, 387–400 (2020).

[35] Masahito Hayashi and Vincent Y. F. Tan. "Minimum rates of approximate sufficient statistics". IEEE Transactions on Information Theory **64**, 875–888 (2018).

[36] Yuxiang Yang, Ge Bai, Giulio Chiribella, and Masahito Hayashi. "Compression for quantum population coding". IEEE Transactions on Information Theory **64**, 4766–4783 (2018).

[37] David Layden, Sisi Zhou, Paola Cappellaro, and Liang Jiang. "Ancilla-free quantum error correction codes for quantum metrology". Phys. Rev. Lett. **122**, 040502 (2019).

[38] Sisi Zhou, Argyris Giannisis Manes, and Liang Jiang. "Achieving metrological limits using ancilla-free quantum error-correcting codes". Phys. Rev. A **109**, 042406 (2024).

[39] Michael J W Hall. "Entropic heisenberg limits and uncertainty relations from the holevo information bound". Journal of Physics A: Mathematical and Theoretical **51**, 364001 (2018).

[40] Majid Hassani, Chiara Macchiavello, and Lorenzo Maccone. "Digital quantum estimation". Phys. Rev. Lett. **119**, 200502 (2017).

[41] Giovanni Chesi, Alberto Riccardi, Roberto Rubboli, Lorenzo Maccone, and Chiara Macchiavello. "Protocol for global multiphase estimation". Phys. Rev. A **108**, 012613 (2023).

[42] Wojciech G'orecki, Xi Lu, Chiara Macchiavello, and Lorenzo Maccone. "Mutual information bounded by fisher information" (2024).

[43] Yuxiang Yang, Renato Renner, and Giulio Chiribella. "Optimal universal

programming of unitary gates". Phys. Rev. Lett. **125**, 210501 (2020).
[44] M. Hayashi. "Quantum information theory: Mathematical foundation". Graduate Texts in Physics. Springer Berlin, Heidelberg. New York, NY (2017). 2nd ed. edition.
[45] Kamil Korzekwa, Zbigniew Puchała, Marco Tomamichel, and Karol Życzkowski. "Encoding classical information into quantum resources". IEEE Transactions on Information Theory **68**, 4518–4530 (2022).
[46] Hiroshi Imai and Masahito Hayashi. "Fourier analytic approach to phase estimation in quantum systems". New Journal of Physics **11**, 043034 (2009).
[47] M. V. Burnashev and A. S. Holevo. "On the reliability function for a quantum communication channel". Problems Inform. Transmission **34**, 042312 (1998).
[48] A. Yu. Kitaev. "Quantum measurements and the abelian stabilizer problem" (1995). arXiv:quant-ph/9511026.
[49] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca. "Quantum algorithms revisited". Proc. R. Soc. Lond. A. **454**, 339–354 (1998).
[50] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman, and G. J. Pryde. "Entanglement-free heisenberg-limited phase estimation". Nature **450**, 393–396 (2007).
[51] Masahito Hayashi. "Fourier analytic approach to quantum estimation of group action". Communications in Mathematical Physics **347**, 3–82 (2016).

[52] D. W. Berry, B. L. Higgins, S. D. Bartlett, M. W. Mitchell, G. J. Pryde, and H. M. Wiseman. "How to perform the most accurate possible phase measurements". Phys. Rev. A **80**, 052114 (2009).
[53] Masahito Hayashi. "Group representation for quantum theory". Springer. (2017).