Generalization capacity of singular models in quantum state estimation

Hiroshi Yano[1*] and Yota Maeda[1,2]

[1*]Quantum Computing Center, Keio University, Hiyoshi 3-14-1, Kohoku, 223-8522, Yokohama, Japan.
[2]Advanced Research Laboratory, Technology Infrastructure Center, Technology Platform, Sony Group Corporation, 1-7-1 Konan, Minato-ku, 108-0075, Tokyo, Japan.

Abstract

Quantum state estimation is an indispensable task in quantum information processing. One solid approach is to use Bayesian inference, quantifying uncertainty in a natural way from experimental data. So far, several quantum state models and prior distributions have been proposed to achieve practical and efficient Bayesian quantum state estimation. However, the statistical behavior of quantum information-theoretic properties of the estimated state, such as quantum relative entropy, has not been clarified yet when the quantum state models are over-parameterized. In the present work, we propose a mathematical theory of singular Bayesian statistics in quantum state estimation based on an algebraic geometrical method. As a main result, we give an asymptotic expansion of a quantum generalization and empirical loss, defined in terms of the quantum relative entropy between the target state and the estimated state. Consequently, we construct an asymptotically unbiased estimator of the quantum generalization loss, *quantum widely applicable information criteria (QWAIC)*. Our results provide a new direction to evaluate the generalization capacity of quantum state models using the quantities introduced in algebraic geometry.

Keywords: quantum state estimation, generalization capacity, quantum information criteria, singular learning theory, algebraic geometry

1 Introduction

Quantum state estimation, or state tomography, [1] is one of the fundamental tasks for advancing quantum technologies. Generally, it requires an exponential amount of measurement data in the system size, which makes it difficult to conduct large-scale state estimation. However, there remains a demand for efficient methods for quantum state estimation. In this regard, Bayesian approaches [2, 3, 4, 5] offer a recipe for practical quantum state estimation taking advantage of prior information.

A parameterization is a key step in quantum state estimation, thus many parametric quantum state models have been proposed so far, including the recent development of neural-network quantum states [6] and quantum Boltzmann machines [7]. Despite these developments, it is difficult to determine which model to use, with little prior knowledge about the target system. This problem has led several studies [8, 9, 10, 11] to address it using a statistical method called *model selection*, for choosing the best model among candidates based on the observed data. In particular, the first-named author proposed quantum information criteria (QIC) [11] based on the novel work by Akaike on AIC [12, 13]. Information criteria represent the bias and variance trade-off and enable the prediction of the performance of an estimated model. QIC evaluates the quality of the estimated quantum state in terms of the quantum relative entropy, a quantum analog of AIC.

However, the proposed methods by these studies assume the regularity condition and often fail to evaluate the model capacity correctly. In classical statistics, a model is said to be *regular* if the map from parameters to probability distributions is one-to-one and its Fisher information matrix is positive definite, in particular, it has an inverse. If otherwise, the model is said to be *singular*. It is known that these singularities in fact appear, for example in mixed Gaussian distributions. In practice, it often appears in the analysis of the neural networks or Bayesian network [14], and in particular, Transformer [15, 16]. For example, the singular learning theory clarifies phase transitions in machine learning [17]. In this vein, we can easily imagine that the regularity condition is not likely to be satisfied when one uses quantum state models with many parameters, such as neural-network quantum states and quantum Boltzmann machines [18, 19]. This motivates us to investigate singular models in quantum state estimation, with the ultimate goal of developing QIC for singular models.

In the present work, we propose a mathematical and Bayesian framework for singular models to estimate quantum states. We build upon the celebrated singular theory for classical Bayesian statistics [14, 20, 21] established by Watanabe to investigate the statistical behavior of quantum information-theoretic properties. For this purpose, we formulate the Bayesian quantum state estimation and define a quantum generalization and empirical loss based on the quantum relative entropy. Our main result describes the asymptotic expansions of quantum generalization and empirical loss based on an algebraic geometrical method. From this description, we propose an asymptotically unbiased estimator, which is called *quantum widely applicable information criteria (QWAIC)*, of the quantum generalization loss. This is a generalization of widely applicable information criteria (WAIC) [14], constructed by Watanabe, for classical singular learning theory. This allows us to evaluate the trade-off between the model adaptability to the observed data and the model capacity.

2 Singular learning theory

We briefly summarize the Bayesian statistics for unknown classical probability distribution [20, 22]. Our goal is to extend this theory to investigate quantum information, discussed in the next section. Assume, given n i.i.d. samples $x^n = \{x_1, ..., x_n\}$ from an unknown probability distribution $q(x)$, one wants to predict $q(x)$ using a pair of a statistical model $p(x|\theta)$ and a prior distribution $\pi(\theta)$, where $\theta \in \Theta \subset \mathbb{R}^d$ and $\theta \sim \pi(\theta)$. Then, the posterior distribution and posterior predictive distribution are immediately defined by

$$p(\theta|x^n) := \frac{1}{p(x^n)}\pi(\theta)\prod_{\alpha=1}^{n} p(x_\alpha|\theta),$$
$$p(x|x^n) := \int_\Theta p(x|\theta)p(\theta|x^n)d\theta$$

where $p(x^n) := \int \pi(\theta)\prod_{\alpha=1}^n p(x_\alpha|\theta)d\theta$ is the marginal likelihood. For this purpose, the classical generalization loss G_n and empirical loss T_n are given by

$$G_n := -\mathbb{E}_X[\log p(X|x^n)], \quad T_n := -\frac{1}{n}\sum_{\alpha=1}^{n}\log p(x_\alpha|x^n).$$

To study the asymptotic behavior of G_n and T_n, let us introduce the Kullback-Leibler divergence and a certain parameter set

$$\mathrm{KL}(q\|p(\cdot|\theta)) := \mathbb{E}_X\left[\log\frac{q(X)}{p(X|\theta)}\right],$$
$$\Theta_0 := \left\{\theta_0 \;\middle|\; \theta_0 = \operatorname*{argmin}_{\theta\in\Theta}\mathrm{KL}(q\|p(\cdot|\theta))\right\}. \quad (1)$$

The function $q(x)$ is said to be *regular* for $p(x|\theta)$ if the following conditions are satisfied:

1. Θ_0 consists of a single element θ_0,
2. the Hessian matrix $\nabla^2\mathrm{KL}(q\|p(\cdot|\theta_0))$ is positive definite, and
3. there is an open neighborhood of θ_0 in Θ.

If otherwise, we call *singular*. Notably, in singular cases, the posterior distribution cannot be approximated by any normal distribution even in the asymptotic limit, and moreover, Θ_0 contains singular points in general, which forces the estimation to be difficult [16].

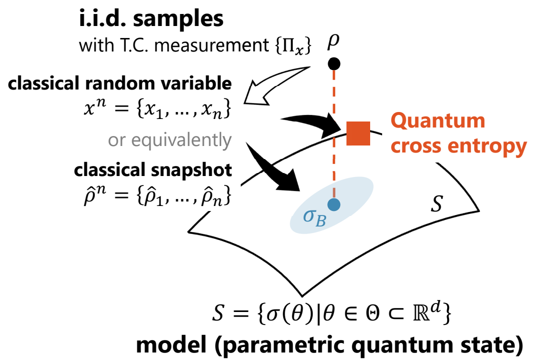

Fig. 1 Our setting in quantum state estimation

Then, even if $q(x)$ is singular for $p(x|\theta)$, the asymptotic behaviors of G_n and T_n are as follows:

$$G_n = -\mathbb{E}_X[\log p(X|\theta_0)] + \frac{1}{n}(\lambda + R - V) + o_P(1/n), \quad (2)$$

$$T_n = -\frac{1}{n}\sum_{\alpha=1}^{n}\log p(x_\alpha|\theta_0) + \frac{1}{n}(\lambda - R - V) + o_P(1/n) \quad (3)$$

where λ, R, and V are defined in [23, 24]. Based on this, Watanabe [14, 20] established the notion of widely applicable information criteria (WAIC) for singular models. It is an asymptotically unbiased estimator of G_n:

$$\mathrm{WAIC} := T_n + \frac{1}{n}\sum_{\alpha=1}^{n}\mathrm{Var}(\log p(x_\alpha|\theta)),$$
$$\mathbb{E}_n[G_n] = \mathbb{E}_n[\mathrm{WAIC}] + o(1/n).$$

Here $\mathrm{Var}(\cdot)$ is the posterior variance, and $\mathbb{E}_n[\cdot]$ in the second equation is the expectation over the sets of n training samples. Hence, it is important to study and analyze the behavior of G_n and T_n for model selection. Note that R and V are characterized by the empirical process of the renormalized log-likelihood functions. The quantity λ is called the real log canonical threshold, which is a well-known birational invariant, represents how bad the singularity of a given $\mathbb{Q}$-divisor is, in algebraic geometry or minimal model program [25, 26]. This measures the effective dimension of the parameter space of a model.

3 Main results

In the present work, we formulate the task of Bayesian quantum state estimation as follows (see Fig. 1). Let ρ be an unknown target state and $x^n = \{x_1, ..., x_n\}$ be a finite number of measurement data obtained through a tomographically complete (T.C.) measurement $\{\Pi_x\}$ with uniform weights on ρ. In other words, the measurement data x^n is a set of the i.i.d. samples from the corresponding true probability distribution $q(x) := \mathrm{Tr}(\Pi_x\rho)$. To predict ρ, one prepares a pair of a parametric quantum state model $\sigma(\theta)$ and a prior distribution $\pi(\theta)$. Let us denote by $p(x|\theta) := \mathrm{Tr}(\Pi_x\sigma(\theta))$ the corresponding probability distribution of the model $\sigma(\theta)$. In addition, we introduce an alternative representation of the measurement data using the classical shadow

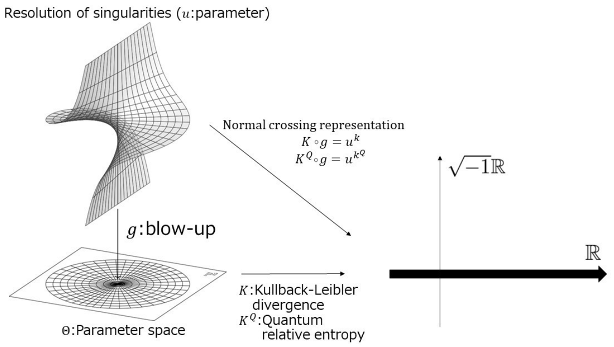

Fig. 2 Resolution of singularities of a parameter space [30]

[27] by $\hat{\rho}^n = \{\hat{\rho}_1, ..., \hat{\rho}_n\}$ where $\hat{\rho}_\alpha$ is called a classical snapshot, corresponding to x_α for each α. With the posterior distribution $p(\theta|x^n)$ defined in the previous section, the posterior predictive quantum state, or simply the Bayesian mean, σ_B is naturally defined by

$$\sigma_B := \int_\Theta \sigma(\theta) p(\theta|x^n) d\theta.$$

Now, the problem is the evaluation of the quantum relative entropy

$$D(\rho||\sigma_B) := \mathrm{Tr}(\rho(\log\rho - \log\sigma_B)),$$

implicitly assuming $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma_B)$. Since the first term of $D(\rho||\sigma_B)$ does not depend on σ_B, it is enough to evaluate the second term, which we call the quantum cross entropy (QCE). Using QCE, we further define the *quantum generalization loss* G_n^{Q} and *training loss* T_n^{Q}:

$$G_n^{\mathrm{Q}} := -\mathrm{Tr}(\rho \log \sigma_B), \tag{4}$$

$$T_n^{\mathrm{Q}} := -\frac{1}{n}\sum_{\alpha=1}^{n} \mathrm{Tr}(\hat{\rho}_\alpha \log \sigma_B). \tag{5}$$

Generally, if our quantum state model $\sigma(\theta)$ is over-parameterized, the regularity condition, defined in the previous section, is no longer satisfied [19], and thus we need to consider $\sigma(\theta)$ as a singular model.

Here we give asymptotic expansions of G_n^{Q} and T_n^{Q} around θ_0 defined in Eq. (1), as Eqs. (2) and (3) in the classical case. Let us write the Kullback-Leibler divergence

$$K(\theta) := \mathrm{KL}(p(\cdot|\theta_0)\|p(\cdot|\theta)) = \mathrm{KL}(\mathrm{Tr}(\Pi_x\sigma(\theta_0))\|\,\mathrm{Tr}(\Pi_x\sigma(\theta)))$$

and the quantum relative entropy $K^{\mathrm{Q}}(\theta) = D(\sigma(\theta_0)\|\sigma(\theta))$. In singular cases, Θ_0 generally contains singular points. Applying Hironaka's theorem on a resolution of singularities [28, 29], we get a proper holomorphic morphism $g : \widetilde{\Theta} \to \Theta$ so that $g^{-1}(\Theta_0)$ is a normal crossing divisor; see Fig. 2 for the case of a nodal curve $y^2 = x^2(x+1)$ [30]. Let u be a parameter of $\widetilde{\Theta}$. Then, for $k, k^{\mathrm{Q}} \in \mathbb{Z}^d$, we have

$$K(\theta) = K(g(u)) = u^k, \quad K^{\mathrm{Q}}(\theta) = K^{\mathrm{Q}}(g(u)) = u^{k^{\mathrm{Q}}}$$

by using multi-indices. Here, we consider the case that there exists a parameter $\theta \in \Theta$ so that $\sigma(\theta) = \rho$ as assumed in AIC. Now, we shall prove the following.

Theorem 1. *Even when $q(x)$ is singular for $p(x|\theta)$, the following asymptotic expansions hold:*

$$G_n^{\mathrm{Q}} = -\mathrm{Tr}(\rho\log\sigma(\theta_0)) + \frac{1}{n}(\lambda^{\mathrm{Q}} + R_1^{\mathrm{Q}}) - \frac{1}{2}V^{\mathrm{Q}} + o_P(1/n),$$

$$T_n^{\mathrm{Q}} = -\frac{1}{n}\sum_{\alpha=1}^{n}\mathrm{Tr}(\hat{\rho}_\alpha\log\sigma(\theta_0)) + \frac{1}{n}(\lambda^{\mathrm{Q}} + R_1^{\mathrm{Q}} - R_2^{\mathrm{Q}}) - \frac{1}{2}V^{\mathrm{Q}} + o_P(1/n)$$

Here $\lambda^{\mathrm{Q}} := \lambda\mathcal{E}[u^{k^{\mathrm{Q}}-k}]$ with the posterior mean $\mathcal{E}[\cdot]$, $V^{\mathrm{Q}} := \mathrm{Tr}(\rho\mathrm{Var}[\log\sigma(g(u))]) = O_P(1/n)$, and R_1^{Q} and R_2^{Q} are characterized by the empirical process of the renormalized log-likelihood functions and its quantum analog using the classical shadow. Note that when $k = k^{\mathrm{Q}}$, the quantity λ^{Q} coincides with λ introduced in [24]. Therefore, our formulas generalize a crucial part of classical singular learning theory to the case of quantum state estimation.

Now, let us define QWAIC as follows.

$$\mathrm{QWAIC} := T_n^{\mathrm{Q}} + \frac{1}{n}\sum_{\alpha=1}^{n}\mathrm{Cov}\left(\log p(x_\alpha|\theta), \mathrm{Tr}(\hat{\rho}_\alpha\log\sigma(\theta))\right).$$

Theorem 2. *Under a certain assumption,* QWAIC *is an asymptotically unbiased estimator of G_n^{Q}. In other words,*

$$\mathbb{E}_n[G_n^{\mathrm{Q}}] = \mathbb{E}_n[\mathrm{QWAIC}] + o(1/n).$$

We here remark that QWAIC can be computed by given samples. The second term of QWAIC is an estimator for R_2^{Q} with the posterior covariance $\mathrm{Cov}(\cdot,\cdot)$. Note that, as discussed in [11] for regular models, QWAIC is expected to incorporate the optimality of the measurement into a model selection criterion, which does not appear in WAIC.

4 Outlook

In Bayesian quantum state estimation for singular models, we provided the asymptotic approximation of the generalization loss and empirical loss in Eqs. (4) and (5) based on the classical singular Bayesian statistics [20, 21]. This leads us to the definition of an asymptotically unbiased estimator QWAIC. Notably, the algebraic geometrical quantity λ^{Q} corresponds to the ratio of quantum Bogoliubov Fisher information to classical Fisher information in regular cases as in [11]. This suggests the connection between algebraic geometry and quantum information geometry. Since many quantum state models, such as quantum Boltzmann machines, have many parameters in practice and are not likely to satisfy the regular condition, finding an efficient estimator for the generalization loss is a highly demanding task in quantum state estimation. we expect that this work provides a new direction in evaluating the model selection of quantum singular models.

Acknowledgements

This work is supported by MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grants No. JPMXS0118067285 and No. JPMXS0120319794.

References

[1] Matteo Paris and Jaroslav Řeháček, editors. *Quantum State Estimation*, volume 649 of *Lecture Notes in Physics*. Springer, Berlin, Heidelberg, 2004.

[2] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12(4):043034, 2010.

[3] Christopher Granade, Joshua Combes, and D. G. Cory. Practical Bayesian tomography. *New Journal of Physics*, 18(3):033024, 2016.

[4] Christopher Granade, Christopher Ferrie, and Steven T. Flammia. Practical adaptive quantum tomography*. *New Journal of Physics*, 19(11):113017, 2017.

[5] Joseph M. Lukens, Kody J. H. Law, Ajay Jasra, and Pavel Lougovski. A practical and efficient approach for Bayesian quantum state estimation. *New Journal of Physics*, 22(6):063038, 2020.

[6] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[7] Mohammad H. Amin, Evgeny Andriyash, Jason Rolfe, Bohdan Kulchytskyy, and Roger Melko. Quantum Boltzmann Machine. *Physical Review X*, 8(2):021050, 2018.

[8] Koji Usami, Yoshihiro Nambu, Yoshiyuki Tsuda, Keiji Matsumoto, and Kazuo Nakamura. Accuracy of quantum-state estimation utilizing Akaike's information criterion. *Physical Review A*, 68(2):022314, 2003.

[9] J. O. S. Yin and S. J. van Enk. Information criteria for efficient quantum state estimation. *Physical Review A*, 83(6):062110, 2011.

[10] Lucia Schwarz and S. J. van Enk. Error models in quantum computation: An application of model selection. *Physical Review A*, 88(3):032318, 2013.

[11] Hiroshi Yano and Naoki Yamamoto. Quantum information criteria for model selection in quantum state estimation. *Journal of Physics A: Mathematical and Theoretical*, 56(40):405301, 2023.

[12] Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Eds. B. N. Petrov and F. Csaki)*, pages 267–281. Akademiai Kiado, Budapest, 1973.

[13] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[14] Sumio Watanabe. *Algebraic geometry and statistical learning theory*, volume 25. Cambridge university press, 2009.

[15] Stan Van Wingerden. Singular theory and alignment. In *Technical AI safety conference*, 2024.

[16] Susan Wei, Daniel Murfet, Mingming Gong, Hui Li, Jesse Gell-Redman, and Thomas Quella. Deep learning is singular, and that's good. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[17] Liam Carroll. *Phase transitions in neural networks*. PhD thesis, MSc Thesis at the University of Melbourne, 2021.

[18] Naoki Yamamoto. On the natural gradient for variational quantum eigensolver. *arXiv preprint arXiv:1909.05074*, 2019.

[19] Martin Larocca, Nathan Ju, Diego García-Martín, Patrick J Coles, and Marco Cerezo. Theory of overparametrization in quantum neural networks. *Nature Computational Science*, 3(6):542–551, 2023.

[20] Sumio Watanabe and Manfred Opper. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12), 2010.

[21] Sumio Watanabe. A widely applicable bayesian information criterion. *The Journal of Machine Learning Research*, 14(1):867–897, 2013.

[22] Sumio Watanabe. Recent advances in algebraic geometry and bayesian statistics. *Information Geometry*, 7(Suppl 1):187–209, 2024.

[23] Sumio Watanabe. *Mathematical theory of Bayesian statistics*. Chapman and Hall/CRC, 2018.

[24] Sumio Watanabe. Mathematical theory of bayesian statistics for unknown information source. *Philosophical Transactions of the Royal Society A*, 381(2247):20220151, 2023.

[25] János Kollár and Shigefumi Mori. *Birational geometry of algebraic varieties. With the collaboration of C. H. Clemens and A. Corti*, volume 134 of *Camb. Tracts Math.* Cambridge: Cambridge University Press, 1998.

[26] Christopher D Hacon, James McKernan, and Chenyang Xu. Acc for log canonical thresholds. *Annals of Mathematics*, pages 523–571, 2014.

[27] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

[28] H. Hironaka. Resolution of singularities of an algebraic variety over a field of characteristic zero. I. *Ann. Math. (2)*, 79:109–203, 1964.

[29] Heisuke Hironaka. Resolution of singularities of an algebraic variety over a field of characteristic zero. II. *Ann. Math. (2)*, 79:205–326, 1964.

[30] Brain Tyrrel. Flatness in algebraic geometry. 2017.