HOKKAIDO UNIVERSITY

Title	Quantum natural gradient without monotonicity
Author(s)	Sasaki, Toi; Miyahara, Hideyuki
Citation	Physical Review A, 110(2), 022439 https://doi.org/10.1103/PhysRevA.110.022439
Issue Date	2024-08-27
Doc URL	https://hdl.handle.net/2115/93510
Rights	©[2024] American Physical Society
Type	journal article
File Information	PhysRevA.110.022439.pdf



Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Quantum natural gradient without monotonicity

Toi Sasaki and Hideyuki Miyahara*

Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Hokkaido 060-0814, Japan

(Received 24 January 2024; revised 29 May 2024; accepted 13 August 2024; published 27 August 2024)

The natural gradient (NG) is an information-geometric optimization method that plays a crucial role, especially in the estimation of parameters for machine learning models like neural networks. To apply NG to quantum systems, the quantum natural gradient (QNG) was introduced and utilized for noisy intermediate-scale devices. Additionally, a mathematically equivalent approach to QNG, known as the stochastic reconfiguration method, has been implemented to enhance the performance of quantum Monte Carlo methods. It is worth noting that these methods are based on the symmetric logarithmic derivative (SLD) metric, which is one of the monotone metrics. So far, monotonicity has been believed to be a guiding principle to construct a geometry in physics. In this paper we propose generalized QNG by removing the condition of monotonicity. Initially, we demonstrate that monotonicity is a crucial condition for conventional QNG to be optimal. Subsequently, we provide analytical and numerical evidence showing that nonmonotone QNG outperforms conventional QNG based on the SLD metric in terms of convergence speed.

DOI: 10.1103/PhysRevA.110.022439

I. INTRODUCTION

Continuous optimization is prevalent in scientific and industrial problem-solving, with a plethora of methods proposed [1]. The natural gradient (NG) stands out as a novel continuous optimization method rooted in information geometry (IG) [2–4]. The field of machine learning has gained significant attention recently, with neural networks (NNs) holding paramount importance [5–7]. NG is frequently employed in the optimization of NNs, demonstrating numerical effectiveness [8].

In the realm of physics, the era of quantum computing has dawned, prompting investigations into the hidden potential of noisy intermediate-scale quantum (NISQ) devices [9]. For NISQ devices, the parameter estimation of quantum circuits is imperative [10–12]. In addition to quantum circuit learning, estimating parameters of quantum states is essential in various contexts, including the variational Monte Carlo method [13], and tensor networks [14]. The quantum natural gradient (QNG), introduced as the quantum counterpart of NG, has recently garnered attention for its promising performance in applications [15,16]. A mathematically equivalent method to QNG, known in condensed matter physics, is the stochastic reconfiguration (SR) algorithm [17–21]. Furthermore, quantum metrics find applications in classical machine learning contexts [22–25]. It is noteworthy that both conventional QNG and the SR method are grounded in the symmetric logarithmic derivative (SLD) metric, which is one of the monotone quantum Fisher metrics and a fundamental concept in quantum information geometry (QIG).

In this paper we present a generalization of QNG based on QIG. Monotonicity implies that information is not gained via completely positive and trace-preserving (CPTP) maps, so it has been believed to be a guiding principle for constructing a geometry in physics. However, from the viewpoint of optimization, it has not been clear whether it is necessary. Initially, we establish that monotonicity is a prerequisite for the SLD metric to optimize the speed of QNG. Subsequently, we theoretically demonstrate that QNG utilizing nonmonotone quantum Fisher metrics exhibits faster performance than its SLD-based counterpart. Furthermore, in practical numerical simulations for both classical and quantum cases, the diagonal approximation of the Fisher metric is often employed to mitigate the computational costs associated with inverting the Fisher metric. We show that even under this approximation, the aforementioned findings remain valid. Lastly, we provide numerical simulations to illustrate that nonmonotone quantum Fisher metrics outperform the SLD metric in terms of speed.

The remainder of this paper is organized as follows. In Secs. II and III we review NG and QNG, respectively. In Sec. IV we explain the concept of monotonicity, which plays an important role in this paper. In Sec. V we present our theoretical results on the speed of generalized QNG. In Sec. VI we provide numerical simulations demonstrating that our algorithm outperforms conventional QNG under certain conditions. Finally, Sec. VII concludes the paper.

II. NATURAL GRADIENT

In the derivation of NG, the Fisher metric and the Kullback-Leibler (KL) divergence play an essential role. We first explain the Fisher metric. Assume that we have $p_\theta(\cdot)$, which is a probability distribution parameterized by $\theta \in \mathbb{R}^n$. We define tangent vectors as $X := \sum_{i=1}^n X^i \partial_i$, where $\partial_i := \frac{\partial}{\partial \theta^i}$ for $i = 1, 2, \ldots, n$. Then, the Fisher metric between X and Y

*Contact author: miyahara@ist.hokudai.ac.jp; hmiyahara512@gmail.com

 ©2024 American Physical Society

is defined as

$$g_{p_\theta(\cdot)}(X,Y) \coloneqq \sum_{x=1}^{N} X^{\mathrm{m}}_{p_\theta(\cdot)}(x) Y^{\mathrm{e}}_{p_\theta(\cdot)}(x), \tag{1}$$

where the e and m representations of X at $p_\theta(\cdot)$ are given, respectively, by

$$X^{\mathrm{m}}_{p_\theta(\cdot)}(\cdot) \coloneqq X p_\theta(\cdot), \tag{2a}$$

$$X^{\mathrm{e}}_{p_\theta(\cdot)}(\cdot) \coloneqq X \ln p_\theta(\cdot). \tag{2b}$$

Next, we turn our attention to the rescaled classical Rényi divergence [26]. When we explain NG only, the KL divergence is sufficient; but the purpose of this paper is to discuss QNG. Then we explain the rescaled classical Rényi divergence, which is a generalized version of the KL divergence. The rescaled classical Rényi divergence is given by

$$D_\alpha(p_{\bar\theta}(\cdot)\|p_\theta(\cdot)) \coloneqq \frac{1}{\alpha(\alpha-1)} \ln \sum_{x=1}^{N} p^\alpha_{\bar\theta}(x) p^{1-\alpha}_\theta(x). \tag{3}$$

Note that "rescaled" means the additional factor $1/\alpha$ in Eq. (3). Furthermore, the rescaled classical Rényi divergence, Eq. (3), in the limit $\alpha \to 1$ corresponds to the KL divergence [26]:

$$\lim_{\alpha\to 1} D_\alpha(p_{\bar\theta}(\cdot)\|p_\theta(\cdot)) = D_{\mathrm{KL}}(p_{\bar\theta}(\cdot)\|p_\theta(\cdot)), \tag{4}$$

where

$$D_{\mathrm{KL}}(p_{\bar\theta}(\cdot)\|p_\theta(\cdot)) \coloneqq \sum_{x=1}^{N} p_{\bar\theta}(x) \ln \frac{p_{\bar\theta}(x)}{p_\theta(x)}. \tag{5}$$

Notably, Eqs. (1) and (3) satisfy the following relationship:

$$g_{p_\theta(\cdot)}(X,Y) = \lim_{\bar\theta\to\theta} \bar X \bar Y D_\alpha(p_{\bar\theta}(\cdot)\|p_\theta(\cdot)), \tag{6}$$

where $\bar X, \bar Y$ are the tangent vectors acting on $p_{\bar\theta}(\cdot)$. The key observation is that the right-hand side of Eq. (6) do not depend on α. As we will see later, the quantum counterpart of Eq. (3) yields different metric contrarily to the classical limit, and this fact will play an important role in QNG.

We here review NG, which is one of the optimization methods for continuous variables motivated by IG [4]. Let us consider the minimization problem of $L(\theta)$:

$$\min_\theta L(\theta). \tag{7}$$

Then, the dynamics of NG is given by

$$\theta_{\tau+1} = \theta_\tau + \Delta\theta_\tau(\epsilon), \tag{8a}$$

$$\Delta\theta_\tau(\epsilon) = \arg\min_{\Delta\theta: D_\alpha(p_{\theta_\tau+\Delta\theta}(\cdot)\|p_{\theta_\tau}(\cdot))\leqslant\epsilon} L(\theta_\tau+\Delta\theta), \tag{8b}$$

where ϵ is a positive number. The key point of Eq. (8) is the constraint in Eq. (8b), since this term does not exist in conventional optimization methods such as the steepest decent method and Newton's method. Note that in this paper we focus on Eq. (3), but we do not need to limit ourselves to Eq. (3). First, we transform Eq. (8b) by using the lowest-order approximation:

$$\Delta\theta_\tau(\epsilon) \approx \arg\min_{\Delta\theta:\frac{1}{2}\Delta\theta^\intercal G_\alpha(\theta_\tau)\Delta\theta\leqslant\epsilon} \nabla_\theta L(\theta_\tau)^\intercal \Delta\theta, \tag{9}$$

where $\nabla_\theta L(\theta) \coloneqq [\partial_1 L(\theta), \partial_2 L(\theta), \ldots, \partial_n L(\theta)]^\intercal$, $\nabla_\theta L(\theta_\tau) \coloneqq \nabla_\theta L(\theta)|_{\theta=\theta_\tau}$, and $G_\alpha(\theta_\tau)$ is the matrix representation of the Fisher metric $g_{p_{\theta_\tau}(\cdot)}(\partial_i,\partial_j)$ induced by $D_\alpha(p_{\theta_\tau+\Delta\theta}(\cdot)\|p_{\theta_\tau}(\cdot))$. To solve Eq. (9), we use the method of Lagrange's multiplier [27]. Then the solution of Eq. (9) is rewritten as

$$\Delta\theta_\tau(\epsilon) = -\sqrt{\frac{2\epsilon}{\nabla_\theta L(\theta_\tau)^\intercal G^{-1}_\alpha(\theta_\tau)\nabla_\theta L(\theta_\tau)}} G^{-1}_\alpha(\theta_\tau)\nabla_\theta L(\theta_\tau). \tag{10}$$

Note that we have added a negative sign to decrease the value of $L(\theta)$ by this update. In a practical usage of NG, we may use the following update rule instead of Eq. (10):

$$\Delta\theta_\tau(\epsilon) = -\eta G^{-1}_\alpha(\theta_\tau)\nabla_\theta L(\theta_\tau), \tag{11}$$

where η is a positive number.

III. QUANTUM NATURAL GRADIENT

To discuss QNG, we explain their quantum counterparts of the Fisher metric, the rescaled classical Rényi divergence. Assume that we have $\hat\rho_\theta$, which is a quantum state parameterized by θ, and its spectral decomposition is given by $\hat\rho_\theta \coloneqq \sum_{i=1}^N p_i|\psi_i\rangle\langle\psi_i|$. We first introduce the quantum extension of the Fisher metric, Eq. (1) [28–31]. To this end, we define the following linear operator $\Delta_{\hat\rho_\theta}$ as follows:

$$\Delta_{\hat\rho_\theta}\hat A \coloneqq \hat\rho_\theta \hat A \hat\rho^{-1}_\theta. \tag{12}$$

And we also introduce the nonlinear transformation of $\Delta_{\hat\rho_\theta}\hat A$:

$$f^{-1}(\Delta_{\hat\rho_\theta})\hat A \coloneqq \sum_{i,j=1}^{N} \frac{\langle\psi_i|\hat A|\psi_j\rangle}{f(p_i/p_j)}|\psi_i\rangle\langle\psi_j|. \tag{13}$$

Using Eq. (13), the quantum counterparts of Eq. (2) at $\hat\rho_\theta$ are defined as follows:

$$\hat X^{\mathrm{m}}_{\hat\rho_\theta} \coloneqq X\hat\rho_\theta, \tag{14a}$$

$$\hat X^{\mathrm{e}}_{\hat\rho_\theta, f(\cdot)} \coloneqq f^{-1}(\Delta_{\hat\rho_\theta})\big([X\hat\rho_\theta]\hat\rho^{-1}_\theta\big), \tag{14b}$$

where $f(\cdot): \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is a operator-monotone function such that

$$f(1) = 1, \tag{15a}$$

$$f(t) = t f(t^{-1}). \tag{15b}$$

Note that $f(\cdot)$ in Eq. (14b) is called the Petz function. Then the quantum Fisher metric is given by

$$g_{\hat\rho_\theta, f(\cdot)}(X,Y) \coloneqq \mathrm{Tr}\big[\hat X^{\mathrm{m}}_{\hat\rho_\theta}\hat Y^{\mathrm{e}}_{\hat\rho_\theta, f(\cdot)}\big]. \tag{16}$$

The key point of the quantum Fisher metric, Eq. (16), is its dependence on the operator-monotone function $f(\cdot)$. Equation (15a) implies the classical limit of the quantum Fisher metric goes to the classical Fisher metric without depending on the choice of $f(\cdot)$, and Eq. (15b) guarantees for the quantum Fisher metric to be real.

We then consider a quantum extension of Eq. (3). The rescaled sandwiched quantum Rényi divergence is defined as follows [32–34]:

$$D_\alpha(\hat\rho_{\bar\theta}\|\hat\rho_\theta) \coloneqq \frac{1}{\alpha(\alpha-1)} \ln \mathrm{Tr}\Big[\Big(\hat\rho_\theta^{\frac{1-\alpha}{2\alpha}}\hat\rho_{\bar\theta}\hat\rho_\theta^{\frac{1-\alpha}{2\alpha}}\Big)^\alpha\Big]. \tag{17}$$

Note that "sandwiched" means $\hat{\rho}_{\bar{\theta}}$ is sandwiched by $\hat{\rho}_\theta$. In Ref. [2], the quantum α divergence, in which $\hat{\rho}_{\bar{\theta}}$ is not sandwiched by $\hat{\rho}_\theta$, is investigated. The rescaled sandwiched Rényi divergence, Eq. (17), becomes the quantum KL divergence [35] in the limit $\alpha \to 1$:

$$\lim_{\alpha\to 1} D_\alpha(\hat{\rho}_{\bar{\theta}}\|\hat{\rho}_\theta) = D_{\mathrm{qKL}}(\hat{\rho}_{\bar{\theta}}\|\hat{\rho}_\theta), \tag{18}$$

where

$$D_{\mathrm{qKL}}(\hat{\rho}_{\bar{\theta}}\|\hat{\rho}_\theta) := \mathrm{Tr}[\hat{\rho}_{\bar{\theta}}(\ln\hat{\rho}_{\bar{\theta}} - \ln\hat{\rho}_\theta)]. \tag{19}$$

Introducing the Petz function given by

$$f_\alpha(t) := (1-\alpha)\frac{1-t^{\frac{1}{\alpha}}}{1-t^{\frac{1-\alpha}{\alpha}}}, \tag{20}$$

Eqs. (16) and (17) satisfy the following relation:

$$g_{\hat{\rho}_\theta, f_\alpha(\cdot)}(X,Y) = \lim_{\bar{\theta}\to\theta} \bar{X}\bar{Y} D_\alpha(\hat{\rho}_{\bar{\theta}}\|\hat{\rho}_\theta), \tag{21}$$

where $\bar{X} := \sum_{i=1}^n X^i\bar{\partial}_i$ and $\bar{Y} := \sum_{i=1}^n Y^i\bar{\partial}_i$.

We revisit the minimization problem, Eq. (7). As the quantum counterpart of Eq. (8), let us consider the minimization problem of $L(\theta)$:

$$\theta_{\tau+1} = \theta_\tau + \Delta\theta_\tau(\epsilon), \tag{22a}$$

$$\Delta\theta_\tau(\epsilon) = \arg\min_{\Delta\theta: D_\alpha(\hat{\rho}_{\theta_\tau+\Delta\theta}\|\hat{\rho}_{\theta_\tau})\leqslant\epsilon} L(\theta_\tau+\Delta\theta). \tag{22b}$$

In Eq. (22b), $\hat{\rho}_\theta$ appears in the constraint instead of $p_\theta(\cdot)$. By using the lowest-order approximation, Eq. (22b) can be transformed into

$$\Delta\theta_\tau(\epsilon) \approx \arg\min_{\Delta\theta:\frac{1}{2}\Delta\theta^{\mathsf{T}}G_\alpha(\theta_\tau)\Delta\theta\leqslant\epsilon} \nabla_\theta L(\theta_\tau)^{\mathsf{T}}\Delta\theta, \tag{23}$$

where $G_\alpha(\theta)$ is the quantum Fisher metric induced by $D_\alpha(\hat{\rho}_{\theta+\Delta\theta}\|\hat{\rho}_\theta)$; that is, it can be rewritten as $[G_\alpha(\theta)]_{i,j} = g_{\hat{\rho}_\theta, f_\alpha(\cdot)}(\partial_i,\partial_j)$ for $i,j=1,2,\ldots,N$, where $[\cdot]_{i,j}$ is the i,jth element of a matrix, because of Eq. (21). To solve Eq. (22), we employ almost the same technique to that used to derive NG. Then we reach the same equation in the classical case, Eqs. (10) and (11), but $G_\alpha^{-1}(\theta_\tau)$ is computed from Eq. (21). Note that we do not need to limit ourselves to Eq. (17), though in this paper we focus on Eq. (17).

IV. MONOTONICITY

Characterizing metrics and divergences, monotonicity is an essential concept [28,36,37]. For example, the monotonicity of the quantum KL divergence is considered to be quite natural since it implies that information is not gained via CPTP maps.

We review some important properties of operator functions [36–39]. We first state the definition of the monotonicity of a function: that is, $f(\cdot)$ is called monotone if and only if

$$\hat{A} \succeq \hat{B} \Rightarrow f(\hat{A}) \succeq f(\hat{B}), \tag{24}$$

where

$$\hat{A} \succeq \hat{B} :\Leftrightarrow \forall|\psi\rangle \in \mathcal{H}, \quad \langle\psi|(\hat{A}-\hat{B})|\psi\rangle \geqslant 0. \tag{25}$$

Metric $g_{\hat{\rho}}(X,Y)$ is called a monotone metric if and only if it satisfies the following relation for an arbitrary CPTP map $\gamma(\cdot)$:

$$g_{\hat{\rho}}(X,X) \geqslant g_{\gamma(\hat{\rho})}(\gamma_*(X),\gamma_*(X)), \tag{26}$$

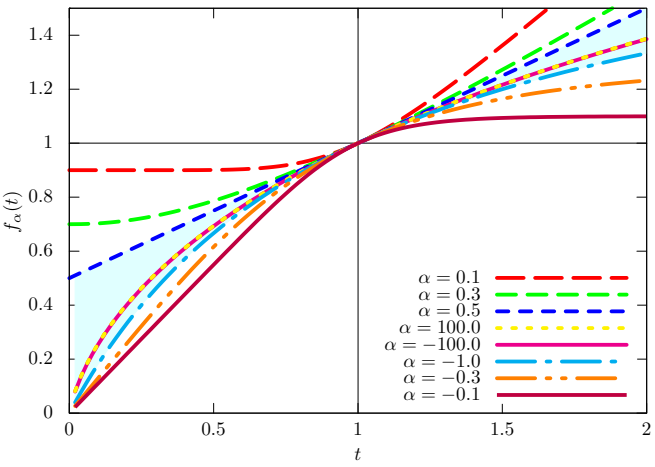


FIG. 1. $f_\alpha(t)$, Eq. (20), with $\alpha = 0.1, 0.3, 0.5, 100.0, -100.0, -1.0, -0.3, -0.1$. Note that $\alpha = 0.5$ and $\alpha = -1.0$ yield the SLD and rRLD metrics, respectively. The regime of the monotone Petz functions, which is between the lines of $\alpha = 0.5$ and $\alpha = -1.0$, is highlighted by light cyan. The lines of $\alpha = 100.0$ and $\alpha = -100.0$ are overlapping with each other.

where $\gamma_*(\cdot)$ is the push forward associated with $\gamma(\cdot)$. Note that Eq. (20) is operator monotone if and only if $\alpha \in (-\infty,-1]\cup[\frac{1}{2},\infty)$ and Eq. (21) is a monotone metric [34].

We define the following partial order on $f(\cdot): \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$:

$$f(\cdot) \preceq \tilde{f}(\cdot) :\Leftrightarrow \forall t \in \mathbb{R}_{\geqslant 0}, \quad f(t) \leqslant \tilde{f}(t). \tag{27}$$

For monotone functions, the following theorem holds [28]:

Theorem 1. $f_{\mathrm{SLD}}(\cdot)$ and $f_{\mathrm{rRLD}}(\cdot)$ are the maximum and minimum elements with respect to the partial order defined in Eq. (27) under the conditions of Eqs. (15) and (24), respectively. That is, we have the following relation for any $f(t)$ such that for Eqs. (15) and (24),

$$f_{\mathrm{rRLD}}(\cdot) \preceq f(\cdot) \preceq f_{\mathrm{SLD}}(\cdot), \tag{28}$$

where $f_{\mathrm{SLD}}(\cdot)$ and $f_{\mathrm{rRLD}}(\cdot)$ are the Petz functions for the SLD and real right logarithmic derivative (rRLD) metrics, respectively, defined as $f_{\mathrm{SLD}}(t) := \frac{1+t}{2}$ and $f_{\mathrm{rRLD}}(t) := \frac{2t}{1+t}$.

We review some important properties of Eq. (20). For simplicity, we substitute $\alpha = \frac{1}{\beta}$ into Eq. (20):

$$f_{\frac{1}{\beta}}(t) = \left(1-\frac{1}{\beta}\right)\frac{1-t^\beta}{1-t^{\beta-1}}. \tag{29}$$

Two important theorems on Eq. (29) are shown in Ref. [34]. The first one is as follows [34]:

Theorem 2. Equation (29) monotonically increases when β increased.

Theorem 2 implies that $f_\alpha(\cdot)$, Eq. (20), is monotonically decreasing with α increasing, but $\alpha = 0$ is singular. Thus, $f_\alpha(\cdot)$, Eq. (20), monotonically increases toward $\alpha \to 0+$. The second one is as follows [34].

Theorem 3. Equation (29) is operator monotone if and only if $\beta \in [-1,2]$.

Theorem 3 means that $f_\alpha(\cdot)$, Eq. (20), is operator monotone for $\alpha \in (\infty,-1]\cup[1/2,\infty)$.

In Fig. 1, we show $f_\alpha(t)$, Eq. (20), for several different α and highlight the regime of the monotone Petz functions by light cyan. Note that $f_\alpha(t)$, Eq. (20), with $\alpha = 1/2, -1$ are the Petz functions for the SLD and rRLD metrics, respectively. Furthermore, we also have $\lim_{\alpha\to\pm\infty} f_\alpha(t) = \frac{t\ln t}{t-1}$ [34].

V. FISHER METRICS AND SPEED OF QNG

We here explain the relation between the quantum Fisher metrics and the speed of QNG. In general, the following theorem holds:

Theorem 4. Let us assume that $f(\cdot)$ and $g_{\hat{\rho}_\theta,f(\cdot)}(\cdot,\cdot)$ are related via Eq. (16). Then, we have

$$f(\cdot) \preceq \tilde{f}(\cdot) \Leftrightarrow g_{\hat{\rho}_\theta,f(\cdot)}(\cdot,\cdot) \succeq g_{\hat{\rho}_\theta,\tilde{f}(\cdot)}(\cdot,\cdot), \tag{30}$$

where

$$g(\cdot,\cdot) \succeq \tilde{g}(\cdot,\cdot) :\Leftrightarrow \forall X, \quad g(X,X) \geqslant \tilde{g}(X,X). \tag{31}$$

The proof of Theorem 4 is shown in Appendix.

We define the following diagonal metric:

$$\bar{g}_{\hat{\rho}_\theta,f(\cdot)}(\partial_i,\partial_j) := \begin{cases} g_{\hat{\rho}_\theta,f(\cdot)}(\partial_i,\partial_j) & (i=j), \\ 0 & (i\neq j). \end{cases} \tag{32}$$

In partial usages, Eq. (32) is of great importance and often used instead of Eq. (16), because the computation of the inverse matrix of Eq. (16) is numerically costly. Similarly to Theorem 4, the following theorem holds:

Theorem 5. The following relation holds for Eq. (32):

$$f(t) \preceq \tilde{f}(\cdot) \Leftrightarrow \bar{g}_{\hat{\rho}_\theta,f(\cdot)}(\cdot,\cdot) \succeq \bar{g}_{\hat{\rho}_\theta,\tilde{f}(\cdot)}(\cdot,\cdot). \tag{33}$$

The proof of Theorem 5 is almost the same with Theorem 4.

We then discuss the dependence of the speed of QNG on the quantum Fisher metrics. The metrics $g(\cdot,\cdot)$ and $\tilde{g}(\cdot,\cdot)$ satisfy the following relation:

$$g(\cdot,\cdot) \succeq \tilde{g}(\cdot,\cdot) \Leftrightarrow g^{-1}(\cdot,\cdot) \preceq \tilde{g}^{-1}(\cdot,\cdot). \tag{34}$$

In the case of Eq. (10), we have

$$L(\theta_{\tau+1}) - L(\theta_\tau) = -\sqrt{2\epsilon\nabla_\theta L(\theta_\tau)^\intercal G_\alpha^{-1}(\theta_\tau)\nabla_\theta L(\theta_\tau)}. \tag{35}$$

And in the case of Eq. (11), we also have

$$L(\theta_{\tau+1}) - L(\theta_\tau) = -\eta\nabla_\theta L(\theta_\tau)^\intercal G_\alpha^{-1}(\theta_\tau)\nabla_\theta L(\theta_\tau). \tag{36}$$

Thus, for both Eqs. (10) and (11), $G_\alpha(\theta_\tau)$ decreases the value of $L(\theta_\tau)$ faster than $G_{\alpha'}(\theta_\tau)$ when $G_\alpha(\theta_\tau) \preceq G_{\alpha'}(\theta_\tau)$.

VI. NUMERICAL SIMULATIONS

We show numerical simulations to support the main claim of this paper. We begin with the setup of numerical simulations. Using $\theta := [\theta_1,\theta_2,\theta_3]^\intercal$, we define the following parameterized quantum state:

$$\hat{\rho}_\theta := \hat{U}(\theta)\hat{\rho}_{\rm ini}\hat{U}^\dagger(\theta), \tag{37}$$

where

$$\hat{U}(\theta) := \hat{R}_z(\theta_3)\hat{R}_y(\theta_2)\hat{R}_z(\theta_1), \tag{38a}$$

$$\hat{R}_z(\phi) := \exp\left(-i\frac{\phi}{2}\hat{\sigma}_z\right), \tag{38b}$$

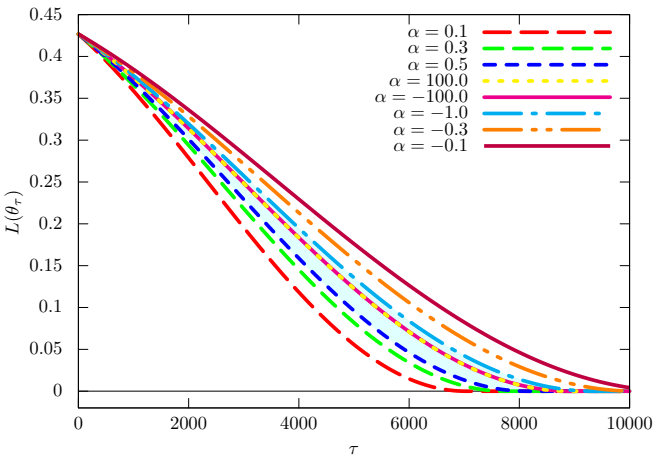


FIG. 2. Cost functions for several α in the case of Eq. (10). We set $\epsilon = 1.0\times10^{-8}$, $\xi = 1.0\times10^{-3}$, and $\delta = 1.0\times10^{-3}$. The regime of the monotone Petz functions, which is between the lines of $\alpha = 0.5$ and $\alpha = -1.0$, is highlighted by light cyan. The lines of $\alpha = 100.0$ and $\alpha = -100.0$ are overlapping with each other.

$$\hat{R}_y(\phi) := \exp\left(-i\frac{\phi}{2}\hat{\sigma}_y\right). \tag{38c}$$

We also define

$$\hat{\rho}_{\rm ini} := \tfrac{1}{2}(\hat{1} + x\hat{\sigma}_x + y\hat{\sigma}_y + z\hat{\sigma}_z), \tag{39}$$

where $\hat{1}$ is the 2×2 identity operator, and $x^2+y^2+z^2 < 1$. Note that $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_z$ are the Pauli operators. Then, we consider the minimization problem, Eq. (7), of the following cost function with respect to θ:

$$L(\theta) := \mathrm{Tr}[(\hat{\rho}_\theta - \hat{\rho}_{\theta_*})^\dagger(\hat{\rho}_\theta - \hat{\rho}_{\theta_*})]. \tag{40}$$

To stabilize the numerical calculations, we conduct the following operations to the density operator and metric:

$$\hat{\rho}_\theta \leftarrow (1-\delta)\hat{\rho}_\theta + \delta\frac{\hat{1}}{N}, \tag{41a}$$

$$G_\alpha(\theta) \leftarrow (1-\xi)G_\alpha(\theta) + \xi I, \tag{41b}$$

where I is the identity matrix whose size is identical to that of $G_\alpha(\theta)$, and δ and η are tiny positive numbers.

In the following calculations, we set $[x,y,z] = [0.5,0.0,0.0]$, $\theta_0 = [\pi/2,\pi/2,\pi/4]^\intercal$, and $\theta_* = [0.0,0.0,0.0]^\intercal$. In Fig. 2 we plot the time evolution of the cost function in the case of Eq. (10) and highlight the regime of the monotone Petz functions by light cyan. In Fig. 3 we show the time evolution of the cost function in the case of Eq. (11). In both cases, QNG with $\alpha \in (0.0, 0.5)$ outperforms SLD-based QNG proposed in Refs. [15,16]. Furthermore, these figures show that the gaps between different values of α become larger in the case of Eq. (11) than Eq. (10). However, this phenomenon depends on the maximum eigenvalue of the quantum Fisher metric; thus, we have to choose either of Eq. (10) or Eq. (11), depending on the problem.

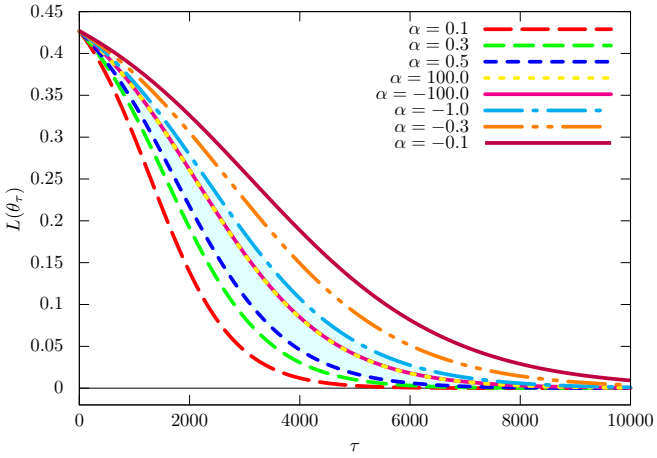


FIG. 3. Cost functions for several α in the case of Eq. (11). We set $\eta = 5.0 \times 10^{-4}$, $\xi = 1.0 \times 10^{-3}$, and $\delta = 1.0 \times 10^{-3}$. The regime of the monotone Petz functions, which is between the lines of $\alpha = 0.5$ and $\alpha = -1.0$, is highlighted by light cyan. The lines of $\alpha = 100.0$ and $\alpha = -100.0$ are overlapping with each other.

VII. CONCLUSIONS

In this paper we have presented a generalization of QNG based on QIG. Notably, the existing literature on QNG relies on the SLD metric, which is not the exclusive definition of the logarithmic derivative of density operators. In QNG, the e representation, specified by the Petz function, aligns with the logarithmic derivative of density operators, and it identifies the quantum Fisher metric. In our work we utilize the nonmonotone quantum Fisher metrics to generalize QNG, successfully enhancing the speed of QNG. Monotonicity has been believed to be a necessary condition to construct a geometry in physics. However, we have clarified that it is not necessary; in fact, it should be eliminated for optimization purposes. Our findings have revealed that mathematical principles considered effective in some fields may not necessarily be advantageous in others.

ACKNOWLEDGMENTS

H. M. was supported by JSPS KAKENHI Grant No. JP23H04489. H. M. thanks A. Fujiwara, L. Wang, Y. Nomura, and S. Daimon for fruitful discussions.

APPENDIX: PROOF OF THEOREM 4

The proof of Theorem 4 is as follows:

Proof. We show Eq. (30). From Eq. (16) we have

$$g_{\hat{\rho}_\theta, f(\cdot)}(X, X) = \sum_{i,j=1}^{N} \frac{1}{p_j f(p_i/p_j)} \langle \psi_j | \hat{X}^{\mathrm{m}}_{\hat{\rho}_\theta} | \psi_i \rangle \langle \psi_i | \hat{X}^{\mathrm{m}}_{\hat{\rho}_\theta} | \psi_j \rangle \tag{A1}$$

$$= \sum_{i,j=1}^{N} \frac{1}{p_j f(p_i/p_j)} |\langle \psi_j | \hat{X}^{\mathrm{m}}_{\hat{\rho}_\theta} | \psi_i \rangle|^2. \tag{A2}$$

Since $f(\cdot) \preceq \tilde{f}(\cdot)$, we have

$$g_{\hat{\rho}_\theta, f(\cdot)}(X, X) = \sum_{i,j=1}^{N} \frac{1}{p_j f(p_i/p_j)} |\langle \psi_j | \hat{X}^{\mathrm{m}}_{\hat{\rho}_\theta} | \psi_i \rangle|^2 \tag{A3}$$

$$\geqslant \sum_{i,j=1}^{N} \frac{1}{p_j \tilde{f}(p_i/p_j)} |\langle \psi_j | \hat{X}^{\mathrm{m}}_{\hat{\rho}_\theta} | \psi_i \rangle|^2 \tag{A4}$$

$$= g_{\hat{\rho}_\theta, \tilde{f}(\cdot)}(X, X). \tag{A5}$$

Thus, Eq. (30) is proved. ■

[1] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering (Springer, New York, 2009).
[2] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry* (American Mathematical Society, Providence, RI, 2000), Vol. 191.
[3] S.-I. Amari, *Information Geometry and Its Applications*, Applied Mathematical Sciences (Springer Japan, Tokyo, 2016).
[4] S.-I. Amari, Neural Comput. **10**, 251 (1998).
[5] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
[6] K. Murphy, *Probabilistic Machine Learning: An Introduction*, Adaptive Computation and Machine Learning Series (MIT Press, Cambridge, MA, 2022).
[7] K. Murphy, *Probabilistic Machine Learning: Advanced Topics*, Adaptive Computation and Machine Learning series (MIT Press, Cambridge, MA, 2023).
[8] J. Martens, J. Mach. Learn. Res. **21**, 5776 (2020).
[9] J. Preskill, Quantum **2**, 79 (2018).
[10] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Phys. Rev. A **98**, 032309 (2018).
[11] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Phys. Rev. A **101**, 032308 (2020).
[12] H. Miyahara and V. Roychowdhury, Sci. Rep. **12**, 19520 (2022).
[13] P. O. Scherer and P. O. Scherer, *Computational Physics: Simulation of Classical and Quantum Systems* (Springer, New York, 2017).
[14] R. Orús, Nat. Rev. Phys. **1**, 538 (2019).
[15] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum **4**, 269 (2020).
[16] B. Koczor and S. C. Benjamin, Phys. Rev. A **106**, 062416 (2022).
[17] S. Sorella and L. Capriotti, Phys. Rev. B **61**, 2599 (2000).
[18] S. Sorella, Phys. Rev. B **64**, 024512 (2001).
[19] G. Mazzola, A. Zen, and S. Sorella, J. Chem. Phys. **137**, 134112 (2012).
[20] C.-Y. Park and M. J. Kastoryano, Phys. Rev. Res. **2**, 023232 (2020).
[21] H. Xie, L. Zhang, and L. Wang, SciPost Phys. **14**, 154 (2023).
[22] A. Lopatnikova, M.-N. Tran, and S. A. Sisson, American Inst. Math. Sci. **6**, 278 (2024).
[23] H. Miyahara, K. Tsumura, and Y. Sughiyama, J. Stat. Mech.: Theory Exp. (2017) 113404.

[24] H. Miyahara and Y. Sughiyama, Phys. Rev. A **98**, 022330 (2018).
[25] H. Miyahara and V. Roychowdhury, Proc. Natl. Acad. Sci. USA **120**, e2212660120 (2023).
[26] T. Van Erven and P. Harremos, IEEE Trans. Inf. Theory **60**, 3797 (2014).
[27] T. Mizohata, T. J. Kobayashi, L.-S. Bouchard, and H. Miyahara, arXiv:2309.10334.
[28] D. Petz, Linear Algebra Appl. **244**, 81 (1996).
[29] D. Petz, *Quantum Information Theory and Quantum Statistics* (Springer Science & Business Media, New York, 2007).
[30] D. Petz and C. Ghinea, *Quantum Probability and Related Topics* (World Scientific, Singapore, 2011), pp. 261–281.
[31] T. Sagawa, *Entropy, Divergence, and Majorization in Classical and Quantum Thermodynamics* (Springer Nature, London, 2022), Vol. 16.
[32] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, J. Math. Phys. **54**, 122203 (2013).
[33] M. M. Wilde, A. Winter, and D. Yang, Commun. Math. Phys. **331**, 593 (2014).
[34] K. Takahashi and A. Fujiwara, J. Phys. A: Math. Theor. **50**, 165301 (2017).
[35] H. Umegaki, *Kodai Mathematical Seminar Reports* (Department of Mathematics, Tokyo Institute of Technology, Tokyo, 1962), Vol. 14, pp. 59–85.
[36] F. Hiai and D. Petz, *Introduction to Matrix Analysis and Applications* (Springer Science & Business Media, New York, 2014).
[37] F. Hiai, Interdiscip. Inf. Sci. **16**, 139 (2010).
[38] R. Bhatia, *Matrix Analysis*, Graduate Texts in Mathematics (Springer, New York, 1996).
[39] R. Horn and C. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, England, 2012).