Fidelity and Fisher information on quantum channels

Haidong Yuan$^{1,3}$ and Chi-Hang Fred Fung$^{2,3}$

$^1$ Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong
$^2$ German Research Center, Huawei Technologies Düsseldorf GmbH, Munich, Germany
$^3$ Authors to whom any correspondence should be addressed.

E-mail: hdyuan@mae.cuhk.edu.hk and fred.fung@huawei.com

Keywords: quantum metrology, Fisher information, quantum information

Abstract

The fidelity function for quantum states has been widely used in quantum information science and frequently arises in the quantification of optimal performances for the estimation and distinguishing of quantum states. A fidelity function for quantum channels is expected to have the same wide applications in quantum information science. In this paper we propose a fidelity function for quantum channels and show that various distance measures on quantum channels can be obtained from this fidelity function; for example, the Bures angle and the Bures distance can be extended to quantum channels via this fidelity function. We then show that the distances between quantum channels lead naturally to a quantum channel Fisher information which quantifies the ultimate precision limit in quantum metrology; the ultimate precision limit can thus be seen as a manifestation of the distances between quantum channels. We also show that the fidelity of quantum channels provides a unified framework for perfect quantum channel discrimination and quantum metrology. In particular, we show that the minimum number of uses needed for perfect channel discrimination is exactly the counterpart of the precision limit in quantum metrology, and various useful lower bounds for the minimum number of uses needed for perfect channel discrimination can be obtained via this connection.

1. Introduction

Fidelity, as a measure of the distinguishability between quantum states [1–3], plays an important role in many areas of quantum information science. For example, it is related to the precision limit in quantum metrology [4], serves as a measure of entanglement preservation through noisy quantum channels [5], and as a measure of entanglement preservation in quantum memory [6]; it has also been used as a characterization method for quantum phase transitions [7], and a criterion for successful transmission in formulating quantum channel capacities [8].

Unlike the fidelity of quantum states, which is defined directly on quantum states, most commonly used measures for the distinguishability of quantum channels are defined indirectly through the effects of the channels on the states. For example, the diamond norm, which is defined as $\|K_1 - K_2\|_0 = \max_{\rho_{SA}} \|K_1 \otimes I_A (\rho_{SA}) - K_2 \otimes I_A (\rho_{SA})\|$ [9–11] (here $\|X\|_0 = \text{Tr} \sqrt{X \dagger X}$, $\rho_{SA}$ denotes a state on system + ancilla, and $I_A$ denotes the identity operator on the ancillary system), is induced by the trace distance on quantum states $\|\rho_1 - \rho_2\|$; another measure for quantum channels which is defined as $\text{arccos} F_{\mu_{\text{max}}}(K_1, K_2) = \text{arccos} \min_{\rho_{SA}} F_2[K_1 \otimes I_A (\rho_{SA}), K_2 \otimes I_A (\rho_{SA})]$ [12, 13], is induced by the fidelity of quantum states $F_2(\rho_1, \rho_2) = \text{Tr} \sqrt{\rho_1^2 \mu_2^2 \mu_1^2}$ . There are also other induced measures which consider the average fidelity defined by an integration over input states over some measures [14–17]. These induced measures through quantum states lack a direct connection to the properties of quantum channels, which severely restricts the insights that can be gained from these measures. A direct measure of quantum channels is expected to provide more insights and is thus highly desired.

In this paper we provide a fidelity function defined directly on quantum channels, and show that this fidelity function for quantum channels, together with the classical fidelity for probability distribution and the fidelity for
quantum states, form a hierarchy of fidelity functions in terms of optimization. This fidelity function for quantum channels also leads to various distance measures defined directly on quantum channels, in particular we show that the Bures angle and the Bures distance can be extended to quantum channels. We then show that the distance between quantum channels leads naturally to a quantum channel Fisher information which quantifies the ultimate precision limit in quantum metrology. We also show that this fidelity function provides a unified framework for perfect quantum channel discrimination and quantum metrology; in particular, we show that the minimum number of uses needed for perfect channel discrimination is exactly the counterpart of the precision limit in quantum metrology, and various useful lower bounds for the minimum number of uses needed for perfect channel discrimination can be obtained via this connection.

2. Fidelity function for quantum channels

We start by defining the fidelity function on unitary channels then extend it to noisy channels.

For an \( m \times m \) unitary matrix \( U \), we denote \( e^{-i \theta_j} \) as the eigenvalues of \( U \), where \( \theta_j \in (−\pi, \pi] \) for \( 1 \leq j \leq m \) and we call \( \theta_j \) the eigen-angles of \( U \). We define (see also [18–20]) \( \| U \|_{max} = \max_{i \leq j \leq m} |\theta_j| \), and \( \| U \|_F \) as the minimum of \( \| e^{i \theta} U \|_{max} \) over equivalent unitary operators with different global phases, i.e.,

\[
\| U \|_F = \min_{\gamma \in \mathbb{R}} \| e^{i \gamma} U \|_{max}.
\]

We then define

\[
C(U) = \begin{cases} 
\| U \|_F, & \text{if } \| U \|_F \leq \frac{\pi}{2}, \\
\frac{\pi}{2}, & \text{if } \| U \|_F > \frac{\pi}{2}.
\end{cases}
\]

Quantitatively \( C(U) \) is equal to the maximal angle that \( U \) can rotate a state away from itself [20–22], i.e.,

\[
\cos(C(U)) = \min_{|\psi\rangle} |\langle \psi | U | \psi \rangle|.
\]

For mixed states it can be written as \( \cos(C(U)) = \min_{\rho} F_\rho(U, \rho U^\dagger) \).

If \( \theta_{\max} = \theta_1 \geq \theta_2 \geq \cdots \geq \theta_m = \theta_{\min} \) are arranged in decreasing order, then \( C(U) = \frac{\theta_{\max} - \theta_{\min}}{2} \) when \( \theta_{\max} - \theta_{\min} \leq \pi \) [20]. We then define \( \Theta_{QC}(U_1, U_2) = C(U_1^\dagger U_2) \) where \( U_1 \) and \( U_2 \) are unitary operators on the same Hilbert space (we can expand the space if they are not the same). It is easy to see that

\[
\Theta_{QC}(U_1, U_2) = \cos(C(U_1^\dagger U_2)) = \min_{\rho} F_\rho(U_1 \rho U_1^\dagger, U_2 \rho U_2^\dagger),
\]

\( \Theta_{QC}(U_1, U_2) \) thus corresponds to the maximal angle between the output states of \( U_1 \) and \( U_2 \) (however we note that the definition of \( \Theta_{QC}(U_1, U_2) \) is independent of the states). We then denote \( F_{QC}(U_1, U_2) = \cos(\Theta_{QC}(U_1, U_2)) \) as the fidelity between \( U_1 \) and \( U_2 \). For unitary channels this is equivalent to the fidelity function proposed previously in [21].

We now generalize this to noisy quantum channels. A general quantum channel \( K \), which maps from \( m_1 \)- to \( m_2 \)-dimensional Hilbert space, can be represented by Kraus operators, \( K(\rho) = \sum_{j=1}^q F_j \rho F_j^\dagger \), where \( \sum_{j=1}^q F_j^\dagger F_j = I \). Equivalently it can also be written as \( K(\rho) = T_{\theta_E}(U_{\theta_E}(\rho \otimes \rho_E) U_{\theta_E}^\dagger) \), where \( \theta_E \) denotes some standard state of the environment, and \( U_{\theta_E} \) is a unitary operator acting on both system and environment, which we call the unitary extension of \( K \).

We define \( \Theta_{QC}(K_s, K_s') = \min_{U_{\theta_{Es1}}, U_{\theta_{Es2}}} \Theta_{QC}(U_{\theta_{Es1}}, U_{\theta_{Es2}}) \) and \( F_{QC}(K_s, K_s') = \cos(\Theta_{QC}(K_s, K_s')) \) when \( U_{\theta_{Es1}} \) and \( U_{\theta_{Es2}} \) are unitary extensions of \( K_s, s \in \{1, 2\} \). In appendix A, we show that the optimization can be taken by fixing one unitary extension and just optimizing over the other unitary extension, i.e.,

\[
\Theta_{QC}(K_1, K_2) = \min_{U_{\theta_{Es1}}} \Theta_{QC}(U_{\theta_{Es1}}, U_{\theta_{Es2}}) = \min_{U_{\theta_{Es2}}} \Theta_{QC}(U_{\theta_{Es1}}, U_{\theta_{Es2}}).
\]

In terms of \( F_{QC}(K_1, K_2) \) it can be written as

\[
F_{QC}(K_1, K_2) = \max_{U_{\theta_{Es1}}} F_{QC}(U_{\theta_{Es1}}, U_{\theta_{Es2}}) = \max_{U_{\theta_{Es2}}} F_{QC}(U_{\theta_{Es1}}, U_{\theta_{Es2}}).
\]

This can be seen as the counterpart of Uhlmann’s purification theorem on quantum channels [23] (however the proof does not use Uhlmann’s purification theorem [24]). In appendix B, we show that \( \Theta_{QC}(K_1, K_2) \) is a metric on quantum channels and can be computed directly from the Kraus operators of \( K_1 \) and \( K_2 \) as [24]

\[
\Theta_{QC}(K_1, K_2) = \arccos \max_{W} \frac{1}{\|W\|_{\infty}^2} \lambda_{\min}(K_W + K_W^\dagger),
\]

where \( \lambda_{\min}(K_W + K_W^\dagger) \) denotes the minimum eigenvalue of \( K_W + K_W^\dagger \) with \( K_W = \sum_{j=1}^q w_j F_j F_j^\dagger \), \( F_1 \) and \( F_2 \) denote the Kraus operators of \( K_1 \) and \( K_2 \) respectively, \( w_j \) denotes the jth entry of a \( q \times q \) matrix \( W \) with \( \| W \|_{\infty} \leq 1 \) where \( \| \cdot \| \) is the operator norm which corresponds to the maximum singular value; here \( W \) arises from the non-uniqueness of the Kraus representations. Thus
\[ F_{\text{QC}}(K_1, K_2) = \max_{\|W\| \leq 1} \frac{1}{2} \lambda_{\min}(K_W + K_W^t). \]  

We emphasize that \( F_{\text{QC}} \) is defined directly on quantum channels without referring to the states. Such direct connection, in contrast to the induced measure, is crucial when applying the fidelity to channel discrimination and quantum metrology, as we will show later. Furthermore, the fidelity can be obtained by substituting \( \|W\| \leq 1 \) which gives
\[ \lambda_{\min}(K_W + K_W^t) = \max_{\|W\| \leq 1} \frac{1}{2} \lambda_{\min}(K_W + K_W^t) \]  
\[ s.t. \quad \begin{bmatrix} I & W \end{bmatrix} \begin{bmatrix} I \end{bmatrix} \succeq 0, \quad K_W + K_W^t - tI \succeq 0. \]  

Analogous to the Bures distance in quantum states \( B_{\text{S}}(\rho_1, \rho_2) = \sqrt{2 - 2 B_{\text{S}}(\rho_1, \rho_2)} \), we can similarly define a Bures distance on quantum channels as \( B_{\text{QC}}(K_1, K_2) = \sqrt{2 - 2 B_{\text{QC}}(K_1, K_2)} \). In appendix A, we prove an intriguing and useful connection between \( B_{\text{QC}}(K_1, K_2) \) and the minimum distances between the Kraus operators of \( K_1 \) and \( K_2 \) as
\[ B_{\text{QC}}^2(K_1, K_2) = \min_{\{F_{1i}, F_{2j}\}} \| \sum_i (F_{1i} - F_{2j})^\dagger (F_{1i} - F_{2j}) \| \]
where \( \{F_{1i}\}, \{F_{2j}\} \) are the sets of all equivalent Kraus representations of \( K_1 \) and \( K_2 \) respectively. This connection is particularly useful in studying the scalings of the distance between quantum channels, as we will show later.

In which sense do we call \( F_{\text{QC}}(K_1, K_2) \) a fidelity function? It turns out that
\[ F_{\text{QC}}(K_1, K_2) = \min_{\rho_{SA}} F_{\text{S}}[K_1 \otimes I_A(\rho_{SA}), K_2 \otimes I_A(\rho_{SA})]. \]  

This equality can be shown by using a result from the supplementary material of [24] which showed that
\[ \min_{\rho_{SA}} F_{\text{S}}[K_1 \otimes I_A(\rho_{SA}), K_2 \otimes I_A(\rho_{SA})] = \max_{\|W\| \leq 1} \frac{1}{2} \lambda_{\min}(K_W + K_W^t), \]
which coincides with equation (6). This gives an operational meaning to \( F_{\text{QC}}(K_1, K_2) \). From this relationship it is also immediately clear that \( F_{\text{QC}}(K_1, K_2) \) is stable, i.e., \( F_{\text{QC}}(K_1 \otimes I, K_2 \otimes I) = F_{\text{QC}}(K_1, K_2) \). We emphasize that although we made connections between \( F_{\text{QC}}(K_1, K_2) \) and the minimum fidelity of the output states, \( F_{\text{QC}}(K_1, K_2) \) is defined directly on quantum channels and does not depend on the states. The definition and the operational meaning of \( F_{\text{QC}}(K_1, K_2) \) play distinct roles in applications, the operational meaning as the minimum fidelity between the output states provides a physical meaning while the direct definition provides a geometrical picture which enables or eases the proofs and computations. This will be demonstrated in the applications. It is analogous to how the fidelity of quantum states is connected to the classical fidelity
\[ F_{\text{S}}(\rho_1, \rho_2) = \min_{\{E_i\}} F_{\text{C}}(p_1, p_2), \]
where \( F_{\text{C}}(p_1, p_2) = \sum_i \sqrt{p_i} \sqrt{p_i} \) denotes the classical fidelity with \( p_{1i} = \text{Tr}(\rho_{1i}) \) and \( p_{2j} = \text{Tr}(\rho_{2j}) \), \( \{E_i\} \) denotes a set of positive operator valued measurements (POVM) [3], here, similarly, the fidelity between quantum states has the operational meaning as the minimum classical fidelity. However, the fidelity between quantum states is defined directly on quantum states, which is independent of the measurements, and such a direct definition has provided numerous insights which would have been hindered with just the classical fidelity.

It is known that the trace distance and the fidelity between quantum states have the following relationships [25]
\[ 1 - F_{\text{S}}(\rho_1, \rho_2) \leq \frac{1}{2} \| \rho_1 - \rho_2 \|_1 \leq \sqrt{1 - F_{\text{S}}^2(\rho_1, \rho_2)}, \]
from which it is straightforward to get the relationships between the diamond norm and the fidelity of quantum channels. This can be obtained by substituting \( \rho_1 = K_1 \otimes I_A(\rho_{SA}) \) and \( \rho_2 = K_2 \otimes I_A(\rho_{SA}) \), then optimizing over \( \rho_{SA} \)
\[ \max_{\rho_{SA}} 1 - F_{\text{S}}[K_1 \otimes I_A(\rho_{SA}), K_2 \otimes I_A(\rho_{SA})] \leq \max_{\rho_{SA}} \frac{1}{2} \| [K_1 \otimes I_A(\rho_{SA}) - K_2 \otimes I_A(\rho_{SA})] \| \]
\[ \leq \max_{\rho_{SA}} \sqrt{1 - F_{\text{S}}^2[K_1 \otimes I_A(\rho_{SA}), K_2 \otimes I_A(\rho_{SA})]}, \]
which gives
\[ 1 - F_{\text{QC}}(K_1, K_2) \leq \frac{1}{2} \| K_1 - K_2 \|_0 \leq \sqrt{1 - F_{\text{QC}}^2(K_1, K_2)}. \]  

Since \( F_{\text{QC}}(K_1, K_2) \) can be computed directly from the Kraus operators, this also provides a way to bound the diamond norm using the Kraus operators.

In [26] the Choi matrices of the quantum channels are used to compute the fidelity between the channels, which corresponds to the fidelity between the output states of two quantum channels when the input state is...
taken as the maximal entangled state. As the maximal entangled state is in general not the optimal input state, the fidelity thus defined does not have the operational meaning as the minimum fidelity of the output states, thus cannot be related to the ultimate precision limit in quantum metrology etc (instead it is related to the precision limit when the probe state is taken as the maximally entangled state).

We give some examples of using the fidelity of quantum channels to quantify the degrading effect of dephasing and spontaneous emission on the dynamics. An ideal unitary phase-ﬂip dynamic is given as $U = \exp(-i\omega t\sigma_z/2)$ where $\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. With the presence of dephasing noises, the dynamics becomes $K_{DP}(\rho) = \sum_{j=1}^2 F_{DPj}\rho_j F_{DPj}^\dagger$ where $F_{DP1} = \left(1 + \frac{i}{2} \sigma_z\right)/2$ and $F_{DP2} = \left(1 - \frac{i}{2} \sigma_z\right)/2$. The fidelity between $U$ and $K_{DP}$ can be computed using equation (6) with $K_W = w_1 U F_{DP1} + w_2 U F_{DP2}$ where $|w_1|^2 + |w_2|^2 \leq 1$. Thus, $(K_W + K_{DP})/2 = \Re\{w_1\} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \Re\{w_2\} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}/2$. Taking the trace at both sides, it can be seen that the trace of $(K_W + K_{DP})/2$ equals $2\Re\{w_1\} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}/2$. This means that $\lambda_{\min}(K_W + K_{DP})/2 \leq \Re\{w_1\} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}/2$, thus $\max_{w_1, w_2} \lambda_{\min}(K_W + K_{DP})/2 \leq \frac{1}{\sqrt{1 + \eta^2}}$, where the equality can be achieved with $w_1 = 1$ and $w_2 = 0$. Therefore, dephasing noise degrades the fidelity to $F_{QC}(U, K_{DP}) = \frac{1}{\sqrt{1 + \eta^2}}$, which does not go to zero as the dephasing rate goes to infinity. The optimal input state achieving this fidelity is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, for which the fidelity between the output states equals to $\frac{1}{\sqrt{1 + \eta^2}}$ and the output states do not become orthogonal when the dephasing rate goes to infinity. For example, the state $\begin{pmatrix} 1 & 0 \end{pmatrix}$ evolves to $\begin{pmatrix} 0 & 1 \end{pmatrix}$ as the dephasing rate goes to infinity, which becomes orthogonal to the output state of the unitary dynamics.

3. A uniﬁed framework for quantum metrology and perfect channel discrimination

Next we demonstrate the applications of the quantum information science, in particular we show how the ﬁdelity provides a uniﬁed platform for the ultimate precision in quantum metrology and the minimum number of uses needed for perfect channel discrimination.

The task of quantum metrology, or quantum parameter estimation in general, is to estimate a parameter $x$ encoded in some channel $K_x$. This can be achieved by preparing a quantum state $\rho_A$ and letting it go through the extended channel $K_x \otimes I_A$ with the output state $\rho_x = K_x \otimes I_A(\rho_A)$. By performing POVM, $\{E_j\}$, on $\rho_x$ one gets the measurement result $y$ with the probability $p(y|x) = \text{Tr}(E_y \rho_x)$. According to the Cramér–Rao bound [27–30], the standard deviation for any unbiased estimator of $x$ is bounded by below by $\sqrt{\delta x} \geq \frac{1}{\text{max}_{y \in \{y\}}[p(y|x)]}$, where $\delta x$ is the standard deviation of the estimation of $x$, $f_c[p(y|x)] = \int_{-\infty}^{\infty} \frac{1}{p(y|x)} \frac{\partial p(y|x)}{\partial x} dy$ is the classical Fisher information and $n$ is the number of times that the procedure is repeated. The classical Fisher information can be further optimized over all POVMs, which gives

$$\delta x \geq \frac{1}{\sqrt{n \text{max}_{y \in \{y\}}[f_c[p(y|x)]]}} = \frac{1}{\sqrt{n h_c(\rho_A)}} \tag{14}$$

where the optimized value $f_c(\rho_A)$ is usually called the quantum Fisher information [4, 27, 28, 31]; here, for clarity, we will call it the quantum state Fisher information.

We first recall established connections between the fidelity functions and the Fisher information. Given $\rho_k$ and its inﬁnitesimal state $\rho_k + dx\sigma_x$ for a given POVM $\{E_j\}$, the classical fidelity between $p(y|x) = \text{Tr}(E_y \rho_x)$ and $p(y|x + dx) = \text{Tr}(E_y \rho_x + dx\sigma_x)$ is given by $F_c[p(y|x), p(y|x + dx)] = \sum_N \sqrt{p(y|x)} \sqrt{p(y|x + dx)}$ which deﬁnes an angle as $\cos \Theta_c[p(y|x), p(y|x + dx)] = F_c[p(y|x), p(y|x + dx)]$. The classical Fisher information
is related to the classical fidelity as \( J_c[p(y|x)] \propto 2 - 2F_c[p(y|x), p(y|x + dx)] \) up to the second order of \( dx \) [4], and this can also be written as

\[
J_c[p(y|x)] = \lim_{dx \to 0} \frac{4\Theta^2_c[p(y|x), p(y|x + dx)]}{dx^2}.
\]  

If we optimize over \( E_x \), the classical fidelity then leads to the fidelity between quantum states as [4]

\[
\min_{\{E_x\}} F_c[\text{Tr}(E_x\rho_x), \text{Tr}(E_x\rho_{x+dx})] = F_S(\rho_x, \rho_{x+dx}),
\]

and the classical Fisher information leads to the quantum state Fisher information \( J_S(\rho_x) = \max_{\{E_x\}} J_c[p(y|x)] \) and up to the second order of \( dx \)

\[
\frac{1}{4} J_S(\rho_x) dx^2 = 2 - 2F_S(\rho_x, \rho_{x+dx}).
\]

If we denote \( \cos \Theta_S(\rho_x, \rho_{x+dx}) = F_S(\rho_x, \rho_{x+dx}) \), then

\[
J_S(\rho_x) = \lim_{dx \to 0} \frac{8[1 - \cos \Theta_S(\rho_x, \rho_{x+dx})]}{dx^2} = \lim_{dx \to 0} \frac{4\Theta^2_S(\rho_x, \rho_{x+dx})}{dx^2}.
\]

The precision can be further improved by optimizing over the probe states, which leads to the ultimate local precision limit of estimating \( x \) from \( K_x \). Intuitively, this ultimate precision limit should be quantified by the distance between \( K_x \) and its infinitesimal neighboring channel \( K_{x+dx} \), in a way analogous to how the distance between quantum states quantifies the precision limit of estimating \( x \) from the state \( \rho_x \) [4]. However, although much progress has been made in calculating the ultimate precision limit [32–40], such a clear physical picture has been missing for more than two decades since Braunstein and Caves’s seminal paper [4]; this is mainly due to the lack of proper tools for quantum channels. Here we show that the fidelity between quantum channels can be used to establish such a physical picture, which also leads naturally to a quantum channel Fisher information.

Further optimizing over the probe states, we have

\[
\max_{\rho_x} \frac{1}{4} J_S(\rho_x) dx^2 = 2 - 2\min_{\rho_{x+dx}} F_S(\rho_x, \rho_{x+dx})
\]

\[
= 2 - 2F_QC(K_x, K_{x+dx})
\]

\[
= B^2_QC(K_x, K_{x+dx}),
\]

and this leads naturally to a quantum channel Fisher information \( J_{QC}(K_x) = \max_{\rho_x} J_S(\rho_x) \) which is similarly related to the distance on quantum channels as

\[
J_{QC}(K_x) = \lim_{dx \to 0} \frac{4B^2_QC(K_x, K_{x+dx})}{dx^2} = \lim_{dx \to 0} \frac{8[1 - \cos \Theta_QC(K_x, K_{x+dx})]}{dx^2} = \lim_{dx \to 0} \frac{4\Theta^2_QC(K_x, K_{x+dx})}{dx^2}.
\]

The quantum channel Fisher information quantifies the ultimate precision limit upon the optimization over the measurements and probe states

\[
\hat{\Delta} \geq \frac{1}{\sqrt{n J_{QC}(K_x)}} = \frac{1}{\sqrt{n} \lim_{dx \to 0} \frac{2b_{QC}(K_x, K_{x+dx})}{|dx|}}.
\]

This connects the precision limit directly to the distance between quantum channels which provides a clear physical picture for the ultimate precision limit. The scaling of the ultimate precision limit can now be seen as a manifestation of the scaling of the distances between quantum channels, as we now show.

Two schemes for multiple uses of quantum channels are usually considered in quantum parameter estimation, the parallel scheme and the sequential scheme, as shown in figure 1. We will show that for both schemes the scaling of the distances between two quantum channels are at most linear, which underlies the scaling for the Heisenberg limit.

For the parallel scheme with \( N \) uses of a channel \( K \), as shown in figure 1(a), the total dynamics can be described by \( K^{\otimes N} \otimes I \). If we denote \( U_{ES} \) as one unitary extension of \( K \), then \( U_{ES}^{\otimes N} \) is a unitary extension of \( K^{\otimes N} \), as shown in figure 2. Given two channels, \( K_1 \) and \( K_2 \), we choose \( U_{ES1} \) and \( U_{ES2} \) as the unitary extension for \( K_1 \) and \( K_2 \), respectively, which satisfies \( \Theta_QC(K_1, K_2) = \Theta_QC(U_{ES1}, U_{ES2}) \). Now as \( U_{ES1}^{\otimes N} \) and \( U_{ES2}^{\otimes N} \) are unitary extensions
of $K_1^{\otimes N}$ and $K_2^{\otimes N}$, respectively, we then have

\[
\Theta_{QC}(K_1^{\otimes N}, K_2^{\otimes N}) \leq \Theta_{QC}(U_{ES1}^{\otimes N}, U_{ES2}^{\otimes N}) \\
= C[(U_{ES1}^{\otimes N}U_{ES2}^{\otimes N})^{\otimes N}] \\
\leq NC(U_{ES1}^t U_{ES2}) \\
= N\Theta_{QC}(K_1, K_2),
\]

(22)

For the sequential scheme (as shown in figure 1(b)), we consider the general case that controls can be inserted between sequential uses of the channels. Any measurements that are used in the control can be substituted by controlled unitaries with ancillary systems, the controls interspersed between the channels can thus be taken as unitaries, which is shown in figure 3. The parallel scheme can be seen as a special case of the sequential scheme by choosing the controls as SWAP gates on the system and different ancillary systems [39]. We show that with $N$ uses of the channel, the distance is still bounded above by $N\Theta_{QC}(K_1, K_2)$.

We present the proof for the case of $N = 2$, and the same line of argument works for general $N$. For $N = 2$, one unitary extension of $U_1 K_1 U_1 K_1$ is $U_1 U_{ES1} U_1 U_{ES1}$, similarly $U_2 U_{ES2} U_1 U_{ES2}$ is a unitary extension of $U_2 K_2 U_1 K_2$, where $U_{ESi}$ denotes a unitary extension of $K_i$, $i = 1, 2$, with $E_i$ as the environment. We can choose $U_{ESi}$ such that $\Theta_{QC}(K_1, K_2) = \Theta_{QC}(U_{ES1}, U_{ES2})$, where all operators are understood as being defined on the whole space so the multiplication makes sense, for example the control $U_1$, which only acts on the system and ancillaries, is understood as $U_1 \otimes I_{E}$, an operator on the whole space including the environment. We then have
i.e., with two uses of the channel, the distance is bounded above by $2\Theta_{\text{QC}}(K_1, K_2)$. With the same line of argument it is easy to show that with $N$ uses of the channel the distance is bounded above by $N\Theta_{\text{QC}}(K_1, K_2)$.

Substituting $K_1$ with $K_{x_1}$ and $K_2$ with $K_{x_2}$, we have

$$\Theta_{\text{QC}}(U_1, K_{x_1}, U_2, K_{x_2}) \leq \Theta_{\text{QC}}(U_1 U_2, K_{x_1}^1, U_2 U_1, K_{x_2}^2, U_1 U_2, K_{x_1}^1, U_2 U_1, K_{x_2}^2, U_1 U_2, K_{x_1}^1, U_2 U_1, K_{x_2}^2)$$

$$= C[U_1 U_2 U_1 U_2 U_1 U_2]\,$$

$$= C(U_1 U_2 U_1 U_2 U_1 U_2) + C[U_1 U_2 U_1 U_2 U_1 U_2]\,$$

$$= 2\Theta_{\text{QC}}(K_{x_1}, K_{x_2}),$$

(23)

The scaling $1/N$ is called the Heisenberg scaling, which, as we showed, is just a manifestation of the fact that the distance between quantum channels grows at most linearly with the number of channels.

For $N$ uses of the channels under the parallel scheme we can also obtain a tighter bound as

$$2 - 2\cos\Theta_{\text{QC}}(K_1^{\otimes N}, K_2^{\otimes N}) \leq N[1 - K_{x_1} - K_{x_2}^N] + N(N - 1)\,\|I - K_{x_1}\|^2,$$

(25)

where $K_{x_1} = \sum_{i=1}^{N} w_i F_i^{x_1}$ as previously defined, and the inequality holds for any $W$ with $\|W\| \leq 1$ (see appendix C). In the asymptotic limit, $N(N - 1)\,\|I - K_{x_1}\|^2$ is the dominating term, and in this case we would like to choose a $W$ minimizing $\|I - K_{x_1}\|$ to get a tighter bound. This can be formulated as semi-definite programming with

$$\min_{\|W\| \leq 1} \|I - K_{x_1}\| = \min_t \,$$

s.t. \quad \begin{pmatrix} I & W^t \\ W & I \end{pmatrix} \succeq 0,$$

$$\begin{pmatrix} t I & (I - K_{x_1})^t \\ (I - K_{x_1}) & t I \end{pmatrix} \succeq 0.$$

(26)

If we let $K_1 = K_x$ and $K_2 = K_{x+d_x}$, then equation (25) provides bounds on the scalings in quantum parameter estimation, which is consistent with the studies in quantum metrology [32, 33, 35, 38, 39] but here it has a more general context (see also [24]).

Next we show how the tools unify quantum parameter estimation and the perfect quantum channel discrimination [21, 41–46].

Given two quantum channels $K_1$ and $K_2$, they can be perfectly discriminated with one use of the channels if and only if there exists a $\rho_{SA}$ such that $K_1 \otimes I_s(\rho_{SA})$ and $K_2 \otimes I_s(\rho_{SA})$ are orthogonal, i.e.,

$$\min_{\rho_{SA}} F_3[K_1 \otimes I_s(\rho_{SA}), K_2 \otimes I_s(\rho_{SA})] = 0,$$

which is the same as $\Theta_{\text{QC}}(K_1, K_2) = \frac{\pi}{2}$. When $K_1$ and $K_2$ cannot be perfectly discriminated with one use of the channel, a finite number of uses may able to achieve the task [45]. This is in contrast to the perfect discrimination of non-orthogonal states which always requires an infinite number of copies. The minimum number of uses needed for perfect channel discrimination should
satisfy $\Theta_{QC}(N\circ K_1, N\circ K_2) = \frac{\pi}{2}$. The perfect channel discrimination is thus determined by the distances between quantum channels, and the scalings of $\Theta_{QC}(N\circ K_1, N\circ K_2)$ obtained before can be used to determine the minimum $N$. For example, from $\Theta_{QC}(N\circ K_1, N\circ K_2) \leq N\Theta_{QC}(K_1, K_2)$ we can obtain a lower bound on $N$ as

$$N \geq \left[ \frac{\pi}{2\Theta_{QC}(K_1, K_2)} \right],$$

(27)

where $[x]$ is the smallest integer not less than $x$. This bound is tighter than existing bounds for noisy channels [43] and for unitary channels it reduces to the formula which is known to be tight [21]. For noisy channels under the parallel scheme we can also substitute $\Theta_{QC}(K_1^\otimes N, K_2^\otimes N) = \frac{\pi}{2}$ into the inequality (25) to get a tighter bound.

The lower bound on minimum $N$ can also be obtained via a connection to quantum metrology. Given two channels $K_1$ and $K_2$, let $K_x, x \in [a, b]$ be a path connecting $K_1$ and $K_2$. With $N$ uses of the channel under the parallel strategy we have

$$\sqrt{\mathcal{I}_{QC}(K_x^\otimes N)} = \lim_{dx \to 0} \frac{2\Theta_{QC}(K_x^\otimes N, K_x^\otimes N)}{|dx|}.$$ From the triangular inequality

$$\Theta_{QC}(K_1^\otimes N, K_2^\otimes N) \leq \int_a^b \lim_{dx \to 0} \frac{\Theta_{QC}(K_x^\otimes N, K_x^\otimes N)}{dx} dx = \frac{1}{2} \int_a^b \sqrt{\mathcal{I}_{QC}(K_x^\otimes N)} dx.$$ (28)

This connects the perfect channel discrimination to the ultimate precision limit. By choosing different paths various useful lower bounds on the minimum number of uses for perfect channel discrimination can be obtained.

For example, take $K_0(\rho) = e^{i\theta} \rho e^{-i\theta}$ and $K_1 = \frac{1 + \eta}{2} \rho + \frac{1 - \eta}{2} \sigma_3$, where $\sigma_1, \sigma_2$ and $\sigma_3$ are Pauli matrices and assume that $\theta = 0.3, \eta = 0.5$. For the parallel strategy the lower bound given by equation (27) is

$$N \geq \left[ \frac{\pi}{2\Theta_{QC}(K_0, K_1)} \right] = 3. $$

If we choose a simple path $K_x = (1 - x)K_0 + xK_1, x \in [0, 1]$, which is a line segment connecting $K_0$ to $K_1$, then with the connection provided by equation (28) we obtain $N \geq 4$. Other paths may be explored to further improve the bound. By using the inequality (25) with $W$ obtained from the semi-definite programming that minimizes $\| I - K_W \|$, we get $N \geq 5$. For any $N$ we can also choose $W$ to minimize $\| 2I - K_W - K_W^\dagger \| + N(N - 1)\| I - K_W \|^2$, and it turns out the minimum $N$ such that $\min_{\| W \| \leq 1} \| 2I - K_W - K_W^\dagger \| + N(N - 1)\| I - K_W \|^2 = 2$ is $6$, thus $N \geq 6$. For comparison, we also explicitly computed the actual distance $\Theta_{QC}(K_0^\otimes N, K_1^\otimes N)$ with the increasing of $N$; it turns out that the minimum $N$ such that $\Theta_{QC}(K_0^\otimes N, K_1^\otimes N) = \frac{\pi}{2}$ is actually $6$. All computations here are done with the CVX package in Matlab [47].

4. Summary

A fidelity function defined directly on quantum channels is provided, which leads to various distance measures defined directly on quantum channels, as well as a quantum channel Fisher information. This forms another hierarchy for fidelity functions and Fisher information, as shown in the table:

$$
\begin{align*}
F_C(p_1, p_2) & \rightarrow F_S(\rho_1, \rho_2) & \rightarrow F_{QC}(K_1, K_2) \\
\Theta_C(p_1, p_2) & \rightarrow \Theta_S(\rho_1, \rho_2) & \rightarrow \Theta_{QC}(K_1, K_2) \\
J_C[p(y|x)] & \rightarrow J_S(\rho_1) & \rightarrow J_{QC}(K_1)
\end{align*}
$$

where $\cos \Theta_S = F_S$ and $J_i = \lim_{dx \to 0} \frac{\Theta_{QC}(K_i, K_i)}{|dx|}, i \in \{ C, S, QC \}$. In this table the functions on quantum states are equal to the optimized value over all measurements of the corresponding functions on probability distribution, and the functions on quantum channels are equal to the optimized value over all probe states of the corresponding functions on quantum states. This framework connects quantitatively the ultimate precision limit and the distance between quantum channels, and provides a clear physical picture for the ultimate precision limit in quantum metrology. It also provides a unified framework for quantum parameter estimation and perfect quantum channel discrimination; with this framework, the progress in one field can be readily transferred to the other field. We expect these tools will find applications in many other fields of quantum information science.
Appendix A. Formula to compute $\Theta_{QC}(K_1, K_2)$

We show that the distance between two quantum channels $\Theta_{QC}(K_1, K_2) = \min_{\{U_{ES1}, U_{ES2}\}} \Theta_{QC}(U_{ES1}, U_{ES2})$ can be computed from the Kraus operators of $K_1$ and $K_2$ as $\Theta_{QC}(K_1, K_2) = \min_{\{U_{ES1}, U_{ES2}\}} \Theta_{QC}(U_{ES1}, U_{ES2}) = \arccos \max_{\|W\| \leq 1} \lambda_{\min}[K_W + K_W^\dagger],$ where $U_{ESi}$ are unitary extensions of $K_i, i \in \{1, 2\}$ and $\lambda_{\min}(K_W + K_W^\dagger)$ denotes the minimum eigenvalue of $K_W + K_W^\dagger$, with $K_W = \sum_{j} w_j F_j^\dagger F_j, F_i$ denotes the Kraus operators of $K_i$ and $K_2, w_i$ denotes the $i$th entry of a $q \times q$ matrix $W$ with $\|W\| \leq 1$ ($\|\cdot\|$ is the operator norm which equals the maximum singular value), and $q$ is the number of the Kraus operators. Furthermore, the minimization on both $U_{ES1}$ and $U_{ES2}$ can be reduced to the minimization of just one

$$\Theta_{QC}(K_1, K_2) = \min_{\{U_{ES1}, U_{ES2}\}} \Theta_{QC}(U_{ES1}, U_{ES2}) = \min_{U_{ES1}} \Theta_{QC}(U_{ES1}, U_{ES2}),$$

(A1)

We start with a general unitary extension for any given channel $K(\rho) = \sum_{j=1}^q F_j \rho F_j^\dagger$ with $\sum_{j=1}^q F_j^\dagger F_j = I$, which maps from a $m_1$- to $m_2$- dimensional Hilbert space,

$$U_{ES} = (W_E \otimes I_{m_2}) \begin{bmatrix} F_1 & * & * & \cdots & * \\ F_2 & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ F_q & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & * \end{bmatrix},$$

(A2)

where $W_E \in U(p)$ only acts on the environment and can be chosen arbitrarily. Here $U(p)$ denotes the set of $p \times p$ unitary operators with $p \geq q$ as $p-q$ zero Kraus operators can be added. Here only the first $m_1$ columns of $U$ are fixed, the freedom of other columns can be represented as

$$U_{ES} = (W_E \otimes I_{m_2}) \begin{bmatrix} F_1 & * & * & \cdots & * \\ F_2 & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ F_q & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & * \end{bmatrix} \begin{bmatrix} I_{m_1} & 0 \\ 0 & V \end{bmatrix},$$

(A3)

where $V$ can be any unitary.

For two channels $K_1$ and $K_2$, with $K_1(\rho) = \sum_{j=1}^q F_{1j} \rho F_{1j}^\dagger$ and $K_2(\rho) = \sum_{j=1}^q F_{2j} \rho F_{2j}^\dagger$, the unitary extensions can be written as

$$U_{ES1} = (W_{E1} \otimes I_{m_2}) \begin{bmatrix} F_{11} & * & * & \cdots & * \\ F_{12} & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ F_{1q} & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & * \end{bmatrix} \begin{bmatrix} I_{m_1} & 0 \\ 0 & V_1 \end{bmatrix},$$

(A4)

$$U_{ES2} = (W_{E2} \otimes I_{m_2}) \begin{bmatrix} F_{21} & * & * & \cdots & * \\ F_{22} & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ F_{2q} & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & * \end{bmatrix} \begin{bmatrix} I_{m_1} & 0 \\ 0 & V_2 \end{bmatrix},$$

(A5)
then

\[ U_{\text{ES1}}^\dagger U_{\text{ES2}} = \begin{bmatrix} I_m & 0 \\ 0 & V_1^\dagger \end{bmatrix} \left[ \begin{array}{cccc} K_W & * & \cdots & * \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{array} \right] \begin{bmatrix} I_m & 0 \\ 0 & V_2 \end{bmatrix}, \]

where \( K_W = \sum_q w_q F_q^1 F_q^2 \), where \( w_q \) is the \( q \)th entry of \( W \), and where \( W \) is the first \( q \times q \) block of \( W_{\text{ES1}} W_{\text{ES2}} \), i.e., \( W_{\text{ES1}} W_{\text{ES2}} = \begin{bmatrix} W & * \\ * & * \end{bmatrix} \). It is easy to see that \( \| W \| \leq 1 \), conversely, for any \( W \) with \( \| W \| \leq 1 \) it can be imbedded as the first \( q \times q \) block of a unitary matrix \([48]\). Thus by varying \( W_{\text{ES1}} \) and \( W_{\text{ES2}} \) we can take \( W \) to be any \( q \times q \) matrix with \( \| W \| \leq 1 \). \( \min_{U_{\text{ES1}}, U_{\text{ES2}}} \Theta_{\text{QC}}(U_{\text{ES1}}, U_{\text{ES2}}) = \min_{U_{\text{ES1}}, U_{\text{ES2}}} C(U_{\text{ES1}}^\dagger U_{\text{ES2}}) \) is now reduced to optimization over \( V_1, V_2 \) and \( W \).

First note that for a fixed \( W \), the first block of \( U_{\text{ES1}}^\dagger U_{\text{ES2}} \) is always \( K_W \), as \( \begin{bmatrix} I_m & 0 \\ 0 & V_1^\dagger \end{bmatrix} \) and \( \begin{bmatrix} I_m & 0 \\ 0 & V_2 \end{bmatrix} \) do not change the first block. It has been shown in \([49]\) that for any unitary that has \( K_W \) as the first block

\[ U = \begin{bmatrix} K_W & * & \cdots & * \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{bmatrix}, \]

\[ \| U \|_{\text{max}} \geq \arccos \left( \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \right), \]

where \( \| U \|_{\text{max}} \) is defined in equation (1) of the main text. Thus we have

\[ \| U_{\text{ES1}}^\dagger U_{\text{ES2}} \|_{\text{max}} \geq \arccos \left( \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \right). \]

What’s more it was also shown that there exists a unitary \( V_2 \) with

\[ U_{V_2} = \begin{bmatrix} K_W & * & \cdots & * \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{bmatrix} \begin{bmatrix} I_m & 0 \\ 0 & V_2 \end{bmatrix}, \]

such that \( \| U_{V_2} \|_{\text{max}} = \arccos \left( \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \right) \) achieves the bound \([49]\). Similarly, the bound can also be achieved by exploring the freedom in rows, i.e., there exists a unitary \( V_1 \),

\[ U_{V_1} = \begin{bmatrix} I_m & 0 \\ 0 & V_1^\dagger \end{bmatrix} \begin{bmatrix} K_W & * & \cdots & * \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{bmatrix}, \]

such that \( \| U_{V_1} \|_{\text{max}} = \arccos \left( \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \right) \). Thus for a fixed \( W, \min_{\| V_1, V_1 \|} \| U_{\text{ES1}}^\dagger U_{\text{ES2}} \|_{\text{max}} = \arccos \left( \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \right) \).

Next we optimize over \( W \). Basically we need to find \( W \) such that \( \arccos \left( \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \right) \) is minimized, which is equivalent to finding \( \max_{\| W \| \leq 1} \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \). Note that the freedom of the global phase from \( \| \cdot \|_{\text{max}} \) to \( \| \cdot \|_{\text{op}} \) (see main text for definitions) has been included in the freedom of \( W \) and since \( \max_{\| W \| \leq 1} \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \geq \frac{1}{2} \lambda_{\min}(K_0 + K_0^\dagger) = 0 \), we have \( \arccos \max_{\| W \| \leq 1} \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \leq \frac{\pi}{2} \). Thus \( \min_{U_{\text{ES1}}, U_{\text{ES2}}} C(U_{\text{ES1}}^\dagger U_{\text{ES2}}) = \arccos \max_{\| W \| \leq 1} \frac{1}{2} \lambda_{\min}(K_W + K_W^\dagger) \), i.e.
\[ \Theta_{QC}(K_1, K_2) = \min_{U_{ES1}, U_{ES2}} \Theta_{QC}(U_{ES1}, U_{ES2}) \]
\[ = \arccos \max_{|W| \leq 1} \frac{1}{2} \lambda_{\text{min}}[K_W + K_W^\dagger]. \quad (A6) \]

It is obvious that the freedom of \( W \) can be achieved by only varying \( W_1 \) or \( W_2 \), thus the equality can be attained by just exploring the freedom of \( V_1 \) and \( V_1^\dagger \), or \( V_2 \) and \( V_2^\dagger \). We then have
\[ \min_{U_{ES1}, U_{ES2}} \Theta_{QC}(U_{ES1}, U_{ES2}) = \min_{U_{ES1}} \Theta_{QC}(U_{ES1}, U_{ES2}) \]
\[ = \arccos \max_{|W| \leq 1} \frac{1}{2} \lambda_{\text{min}}[K_W + K_W^\dagger]. \quad (A7) \]

Next we show that this distance measure has a connection to the minimum distance between equivalent Kraus operators. Given two quantum channels, \( K_1(\rho_\ell) = \sum_{i=1}^{d_1} f_{i1} \rho_{\ell} f_{i1}^\dagger \), and \( K_2(\rho_\ell) = \sum_{i=1}^{d_2} f_{i2} \rho_{\ell} f_{i2}^\dagger \) (zero Kraus operators can be appended if the number of the Kraus operators are not the same), by appending additional \( p - q \) zero Kraus operators, we have the Kraus operators for \( K_1 \) and \( K_2 \) as \( \{ f_{i1}, f_{i2}, \cdots, f_{ip}, 0, \cdots, 0 \} \) and \( \{ f_{i1}, f_{i2}, \cdots, f_{ip}, 0, \cdots, 0 \} \), respectively. Equivalent Kraus operators for \( K_1 \) and \( K_2 \) can be represented as \( \tilde{F}_{il} = \sum_k u_{ik} f_{ik} \) and \( \tilde{F}_{2l} = \sum_k v_{ik} f_{ik} \) where \( u_{ik} \) and \( v_{ik} \) are entries of \( U_1 \), \( V \in U(p) \), respectively, where \( 1 \leq i \leq p \). Then
\[ \min_{\{F_{i1}\}, \{F_{i2}\}} \left\| \sum_{l=1}^{p} (\tilde{F}_{il} - \tilde{F}_{2l}) (\tilde{F}_{il} - \tilde{F}_{2l})^\dagger \right\| = \min_{\{F_{i1}\}, \{F_{i2}\}} \left\| 2I - \sum_{l=1}^{p} f_{i1} f_{i1}^\dagger + f_{i2} f_{i2}^\dagger \right\| \]
\[ = \max_{|W| \leq 1} \{ 2 - \lambda_{\text{min}}(K_W + K_W^\dagger) \}, \quad (A8) \]

where \( K_W = \sum_{j=1}^{q} w_j f_{i1} f_{i2}^\dagger \) and \( w_j \) is the \( j \)th entry of \( W \), which is the first \( q \times q \) block of \( U^\dagger V \) and can be any \( q \times q \) matrix with \( ||W|| \leq 1 \) by varying \( U \) and \( V \), i.e., by varying the equivalent representations of \( K_1 \) and \( K_2 \). Thus
\[ \min_{\{F_{i1}\}, \{F_{i2}\}} \left\| \sum_{l=1}^{p} (\tilde{F}_{il} - \tilde{F}_{2l}) (\tilde{F}_{il} - \tilde{F}_{2l})^\dagger \right\| = \min_{|W| \leq 1} \{ 2 - \lambda_{\text{min}}(K_W + K_W^\dagger) \} \]
\[ = 2 - \max_{|W| \leq 1} \lambda_{\text{min}}(K_W + K_W^\dagger) \]
\[ = 2 - 2 \cos \Theta_{QC}(K_1, K_2), \quad (A9) \]

and we then have
\[ B_{QC}^2(K_1, K_2) = 2 - 2 \cos \Theta_{QC}(K_1, K_2) = \min_{\{F_{i1}\}, \{F_{i2}\}} \left\| \sum_{l=1}^{p} (\tilde{F}_{il} - \tilde{F}_{2l}) (\tilde{F}_{il} - \tilde{F}_{2l})^\dagger \right\|. \quad (A10) \]

**Appendix B. \( \Theta_{QC}(K_1, K_2) \) defines a metric on quantum channels**

We show that \( \Theta_{QC}(K_1, K_2) \) defines a metric on quantum channels.

First we show that \( \Theta_{QC}(U_1, U_2) = C(U_1^\dagger U_2) \), where \( C \) is defined in the main text, is a metric on unitary channels.

We start by listing some useful properties of \( C(U) \):
\[ C(U^\dagger UV) = C(U) \]
\[ C(U_1 \otimes U_2) \leq C(U_1) + C(U_2) \]
\[ C(U_1 U_2) \leq C(U_1) + C(U_2) \]
\[ (B1) \]

where \( V \) is any unitary operator. The first equality is obvious from the definition. The second inequality can be easily verified using the formula \( C(U) = \frac{\theta_{\text{max}} - \theta_{\text{min}}}{2} \) when \( \theta_{\text{max}} - \theta_{\text{min}} \leq \pi \); the equality is saturated when \( C(U_1) + C(U_2) \leq \frac{\pi}{2} \); proof of the third inequality can be found in [50, 51].
It is obvious that $\Theta_{QC}(U, U) = 0$ and $\Theta_{QC}(U_1, U_2) = \Theta_{QC}(U_2, U_1) > 0$ if $U_1 \neq U_2$. And since

$$\Theta_{QC}(U_1, U_2) = C(U_1^\dagger U_2)$$

$$\leq C(U_1^\dagger U_2) + C(U_2^\dagger U_3)$$

$$\leq \Theta_{QC}(U_1, U_3) + \Theta_{QC}(U_3, U_2)$$

where for the inequality we have used the property $C(U_1 U_2) \leq C(U_1) + C(U_2)$. This shows that $\Theta_{QC}(U_1, U_2)$ is a metric on unitary operators.

For two general channels, $\Theta_{QC}(K_1, K_2) = \min_{U_{ES1}} \Theta_{QC}(U_{ES1}, U_{ES2}) = \min_{U_{ES1}} \Theta_{QC}(U_{ES1}, U_{ES2})$ where $U_{ES1}$ and $U_{ES2}$ are unitary extensions for $K_1$ and $K_2$, respectively. It is easy to see that $\Theta_{QC}(K_1, K_2) = \Theta_{QC}(K_1, K_2) \geq 0$ and the equality is saturated only when $K_1 = K_2$. We show that $\Theta_{QC}$ also satisfies the triangular inequality as

$$\Theta_{QC}(K_1, K_3) = \min_{U_{ES1}} \Theta_{QC}(U_{ES1}, U_{ES3})$$

$$\leq \min_{U_{ES1}} C(U_{ES1}, U_{ES3})$$

$$\leq \min_{U_{ES1}} C(U_{ES1}, U_{ES2}) + \Theta_{QC}(K_2, K_3)$$

The last inequality is valid for any $U_{ES2}$. We can choose the $U_{ES2}$ which minimizes $C(U_{ES1}, U_{ES3})$, thus

$$\Theta_{QC}(K_1, K_3) \leq \min_{U_{ES1}} \Theta_{QC}(U_{ES1}, U_{ES3}) + \Theta_{QC}(K_2, K_3)$$

$$\Theta_{QC}(K_1, K_3) = \Theta_{QC}(K_1, K_2) + \Theta_{QC}(K_2, K_3).$$

$\Theta_{QC}(K_1, K_2)$ thus defines a metric on the space of quantum channels.

Appendix C. Upper bound of the distance with $N$ parallel channels

Given two quantum channels, $K_1(\rho) = \sum_{i=1}^N F_i \rho F_i^\dagger$ and $K_2(\rho) = \sum_{i=1}^N F_i \rho F_i^\dagger$, by appending $p - q$ zero Kraus operators, we have the Kraus operators for $K_1$ and $K_2$ as $\{F_{1i}, F_{12}, \cdots, F_{1q}, 0, \cdots, 0\}$ and $\{F_{2i}, F_{22}, \cdots, F_{2p}, 0, \cdots, 0\}$ respectively. All the equivalent Kraus operators for $K_1$ and $K_2$ can be represented as $\bar{F}_{1i} = \sum_k u_{ik} F_{1k}$ and $\bar{F}_{2i} = \sum_k v_{ik} F_{2k}$ where $u_{ik}$ and $v_{ik}$ are entries of $U$, $V \in U(p)$ respectively, where $1 \leq i \leq p$.

With $N$ channels in parallel, one representation of the Kraus operators for $K_1^{\otimes N}$ can be written as $\bar{F}_{i1,1,\cdots,1} = \bar{F}_{11}^{(1)} \otimes \bar{F}_{11}^{(2)} \otimes \cdots \otimes \bar{F}_{11}^{(N)}$, and similarly for $K_2^{\otimes N}$ we have $\bar{F}_{i1,2,\cdots,2} = \bar{F}_{21}^{(1)} \otimes \bar{F}_{21}^{(2)} \otimes \cdots \otimes \bar{F}_{21}^{(N)}$, where $\bar{F}_{1k}^{(l)} = \sum_{i=1}^p u_{ik} F_{1k}$ is one Kraus operator of the $l$th channel of $K_1^{\otimes N}$. Similarly, $\bar{F}_{2k}^{(l)} = \sum_{i=1}^p v_{ik} F_{2k}$ is one Kraus operator of the $l$th channel of $K_2^{\otimes N}$. As $\{\bar{F}_{11,1,\cdots,1}\}$ and $\{\bar{F}_{21,2,\cdots,2}\}$ are just one particular Kraus representation of $K_1^{\otimes N}$ and $K_2^{\otimes N}$, respectively, we then have

$$2 - 2 \cos \Theta_{QC}(K_1^{\otimes N}, K_2^{\otimes N}) \leq \left\| \sum_{i,j=1}^{1,2} (\bar{F}_{i1,1,\cdots,1} - \bar{F}_{i2,2,\cdots,2})^\dagger \right\|$$

$$\times (\bar{F}_{i1,1,\cdots,1} - \bar{F}_{i2,2,\cdots,2})$$

since

$$\bar{F}_{i1,1,\cdots,1} - \bar{F}_{i2,2,\cdots,2} = \bar{F}_{11}^{(1)} \otimes \cdots \otimes \bar{F}_{i1}^{(l)} \otimes \cdots \otimes \bar{F}_{11}^{(N)} - \bar{F}_{21}^{(1)} \otimes \cdots \otimes \bar{F}_{i2}^{(l)} \otimes \cdots \otimes \bar{F}_{21}^{(N)}$$

$$= (\bar{F}_{i1}^{(1)} - \bar{F}_{i2}^{(1)}) \otimes \bar{F}_{11}^{(2)} \otimes \cdots \otimes \bar{F}_{i1}^{(N)} + \bar{F}_{i2}^{(1)} \otimes (\bar{F}_{i2}^{(2)} \otimes \cdots \otimes \bar{F}_{i1}^{(N)} - \bar{F}_{i1}^{(2)} \otimes \cdots \otimes \bar{F}_{i1}^{(N)})$$

$$+ \bar{F}_{i2}^{(2)} \otimes \cdots \otimes (\bar{F}_{i1}^{(N)} - \bar{F}_{i2}^{(N)}),$$

and by induction it is then easy to get that

$$\bar{F}_{i1,1,\cdots,1} - \bar{F}_{i2,2,\cdots,2} = (\bar{F}_{i1}^{(1)} - \bar{F}_{i2}^{(1)}) \otimes \bar{F}_{11}^{(2)} \otimes \cdots \otimes \bar{F}_{i1}^{(N)}$$

$$+ \bar{F}_{i2}^{(1)} \otimes (\bar{F}_{i2}^{(2)} - \bar{F}_{i2}^{(1)}) \otimes \bar{F}_{i2}^{(3)} \otimes \cdots \otimes \bar{F}_{i1}^{(N)}$$

$$+ \bar{F}_{i2}^{(2)} \otimes \bar{F}_{i2}^{(3)} \otimes (\bar{F}_{i1}^{(3)} - \bar{F}_{i2}^{(3)}) \otimes \cdots \otimes \bar{F}_{i1}^{(N)}$$

$$+ \cdots + \bar{F}_{i2}^{(l-1)} \otimes \bar{F}_{i2}^{(l)} \otimes \cdots \otimes (\bar{F}_{i1}^{(N)} - \bar{F}_{i2}^{(N)}).$$
Thus
\[
\sum_{i_1, i_2, \ldots, i_N} (\tilde{F}_{i_1, i_2, \ldots, i_N} - \tilde{F}_{2i_1, 2i_2, \ldots, 2i_N}) (\tilde{F}_{i_1, i_2, \ldots, i_N} - \tilde{F}_{2i_1, 2i_2, \ldots, 2i_N}) = \sum_{i_1, i_2, \ldots, i_N} I \otimes I \otimes \cdots \otimes \left[ \sum_{i} (\tilde{F}_{i_1}^{(i)} - \tilde{F}_{2i_1}^{(i)}) (\tilde{F}_{i_1}^{(i)} - \tilde{F}_{2i_1}^{(i)}) \right] \otimes I \otimes \cdots \otimes I
\]
\[
+ \sum_{i_1, i_2, \ldots, i_N} I \otimes I \otimes \cdots \otimes \left[ \sum_{i} (\tilde{F}_{i_1}^{(i)} - \tilde{F}_{2i_1}^{(i)}) (\tilde{F}_{i_1}^{(i)} - \tilde{F}_{2i_1}^{(i)}) \right] \otimes \cdots \otimes \left[ \sum_{i} (\tilde{F}_{i_1}^{(i)} - \tilde{F}_{2i_1}^{(i)}) (\tilde{F}_{i_1}^{(i)} - \tilde{F}_{2i_1}^{(i)}) \right] \otimes \cdots \otimes I + h.c.
\]

Here again $W$ is the first $q \times q$ block of $U^\dagger V$ and $K_W = \sum_{i=1}^q w_i \tilde{F}_i \tilde{F}_i$ with $w_i$ being the $i$th entry of $W$. We then have
\[
2 - 2 \cos \Theta_{QC}(K_{11}^{(N)}, K_{22}^{(N)}) \leq \left\| \sum_{i_1, i_2, \ldots, i_N} (\tilde{F}_{i_1, i_2, \ldots, i_N} - \tilde{F}_{2i_1, 2i_2, \ldots, 2i_N}) (\tilde{F}_{i_1, i_2, \ldots, i_N} - \tilde{F}_{2i_1, 2i_2, \ldots, 2i_N}) \right\|
\]
\[
\leq N\|2I - K_W - \hat{K}_W^\dagger \| + N(N - 1)\|I - K_W\|^2.
\]

References

[1] Fuchs C A and Caves C M 1994 Phys. Rev. Lett. 73 3047
[2] Jozsa R 1994 J. Mod. Opt. 41 2315
[3] Fuchs C A 1996 arXiv:quant-ph/9601020
[4] Braunstein S L and Caves C M 1995 Phys. Rev. Lett. 74 1339
[5] Schumacher B 1996 Phys. Rev. A 54 2614
[6] Surmacz K, Nunn J, Waldermann F C, Wang Z, Walmsley I A and Jaksch D 2006 Phys. Rev. A 74 050302
[7] Gu S-J 2010 J. Mod. Phys. B 24 4371
[8] Barnum H, Knill E and Nielsen M A 2000 IEEE Trans. Inf. Theor. 46 1317
[9] Bitade A 1997 Russian Mathematical Surveys 52 1191
[10] Bitade A Y, Shen A H and Vylayi M N 2002 Classical and Quantum Computation (Bostons, MA: American Mathematical Society)
[11] Watrous J 2009 arXiv:0901.4709
[12] Gilchrist A, Langford N K and Nielsen M A 2005 Phys. Rev. A 71 062310
[13] Belavkin V P, D’Ariano G M and Raginsky M 2005 J. Math. Phys. 46 062106
[14] Cabrera R and Baysia W E 2007 Phys. Lett. A 368 25–8
[15] Bagan E, Baig M and D’Ariano G M 2006 Phys. Rev. A 87 042140
[16] Pedersen L H, Moller N M and Mølmer K 2007 Phys. Rev. A 87 47–51
[17] Emerson J, Alicki R and Życzkowski K 2005 J. Opt. B: Quantum Semiclass. Opt 7 S347–52
[18] Chau H F 2011 Quant. Inf. Comput. 11 0721
[19] Fung C-H F and Chau H F 2013 Phys. Rev. A 88 012307
[20] Fung C-H F and Chau H F 2014 Phys. Rev. A 90 022333
[21] Acin A 2001 Phys. Rev. Lett. 87 177901
[22] D’Ariano G M, Lo Presti P and Paris M G A 2001 Phys. Rev. Lett. 87 270404
[23] Uhlmann A 1976 Rep. Math. Phys. 9 273–9
[24] Yuan H and Fung C-H F 2017 npj Quant. Inf. 3 14
[25] Fuchs C A and van de Graaf J 1999 IEEE Trans. Inf. Theory 45 1216
[26] Raginsky M 2001 Phys. Lett. A 290 11–8
[27] Helstrom C W 1976 Quantum Detection and Estimation Theory (NewYork: Academic Press)
[28] Holevo A S 1982 Probabilistic and Statistical Aspects of Quantum Theory (New York: North-Holland)
[29] Cramer H 1946 Mathematical Methods of Statistics (Princeton, NJ: Princeton University Press)
[30] Rao C R B 1945 Calcutta Math. Soc. 37 81
[31] Braunstein S L, Caves C M and Milburn G J 1996 Ann. Phys., NY 247 135
[32] Fujiwara A and Imam H 2008 J. Phys. A: Math. Theor. 41 253040
[33] Escher B M, de Matos Filho R I and Davideovich L 2011 Nat. Phys. 7 406
[34] Tsang M 2013 New J. Phys. 15 073005
[35] Demkowicz-Dobrzański R, Kolodyński J and Guta M 2012 Nat. Commun. 3 1063
[36] Knyszyński V N and Durkin G A 2011 Phys. Rev. A 83 021804
[37] Knyszyński S, Chen E and Durkin G 2014 arXiv:1402.0495
[38] Kolodyński J and Demkowicz-Dobrzański R 2013 New J. Phys. 15 073043
[39] Demkowicz-Dobrzański R and Maccone L 2014 Phys. Rev. Lett. 113 250801
[40] Alipour S, Mehboodi M and Rezakhani A T 2014 Phys. Rev. Lett. 112 120405
[41] Duan R, Feng Y and Ying M 2007 Phys. Rev. Lett. 98 100503
[42] Duan R, Feng Y and Ying M 2008 Phys. Rev. Lett. 100 020503
[43] Lu C, Chen J and Duan R 2012 Quantum Info. Comput. 12 138
[44] Chiribella G, D’Ariano G M and Perinotti P 2008 Phys. Rev. Lett. 101 180501
[45] Duan R, Feng Y and Ying M 2009 Phys. Rev. Lett. 103 210501
[46] Harrow A W, Hassidim A, Leung D W and Watrous J 2010 Phys. Rev. A 81 032339
[47] Grant M and Boyd S 2011 CVX: Matlab Software for Disciplined Convex Programming http://cvxr.com/cvx/
[48] Choi M D and Li C K 2001 J. Op. Theor. 46 435
[49] Fung C-H F, Chau H F, Li C K and Sue N S 2015 Quant. Inf. Comp. 15 0685–93
[50] Childs A, Preskill J and Renes J 2000 Quantum information and precision measurement J. Mod. Opt. 47 155–76
[51] Yuan H and Fung C-H F 2015 Phys. Rev. Lett. 115 110401