Allowed region and optimal measurement for information versus disturbance in quantum measurements

Hiroaki Terashima

Department of Physics, Faculty of Education, Gunma University, Maebashi, Gunma 371-8510, Japan

Abstract

We present graphs of information versus disturbance for general quantum measurements of completely unknown states. Each piece of information and disturbance is quantified by two measures: (i) the Shannon entropy and estimation fidelity for the information and (ii) the operation fidelity and physical reversibility for the disturbance. These measures are calculated for a single outcome and are plotted on four types of information–disturbance planes to show their allowed regions. In addition, we discuss the graphs of these metrics averaged over all possible outcomes and the optimal measurements when saturating the upper bounds on the information for a given disturbance. The results considerably broaden the perspective of trade-offs between information and disturbances in quantum measurements.

PACS: 03.65.Ta, 03.67.-a
Keywords: quantum measurement, quantum information

1 Introduction

In quantum theory, a measurement that provides information about a system inevitably disturbs the state of the system, unless the original state is a classical mixture of the eigenstates of an observable. This feature is not only of great interest to the foundations of quantum mechanics but also plays an important role in quantum information processing and communication [1], such as in quantum cryptography [2–5]. As a result, the relationship between information and disturbances has been the subject of numerous studies [6–22] over many years. Most studies have only discussed the disturbance in terms of the size of the state change. However, the disturbance can also be discussed in terms of the reversibility of the state change [23,26] because the state change
can be recovered with a nonzero probability of success if the measurement is physically reversible \cite{27,29}.

Intuitively, if a measurement provides more information about a system, the measurement changes the state of the system by a greater degree and the change becomes more irreversible. To show this trade-off, various inequalities have been derived using different formulations. For example, Banaszek \cite{7} derived an inequality between the amount of information gain and the size of the state change using two fidelities, and Cheong and Lee \cite{25} derived an inequality between the amount of information gain and the reversibility of the state change using the fidelity and reversal probability. These inequalities have been verified \cite{30–33} in single-photon experiments.

In this paper, we present graphs of information versus disturbance for general quantum measurements of a $d$-level system in a completely unknown state. The information is quantified by the Shannon entropy \cite{6} and the estimation fidelity \cite{7}, whereas the disturbance is quantified by the operation fidelity \cite{7} and the physical reversibility \cite{34}. These metrics are calculated for a single outcome using the general formulas derived in Ref. \cite{26} and are plotted on four types of information–disturbance planes to show the allowed regions. Moreover, we show the allowed regions for these metrics averaged over all possible outcomes via an analogy with the center of mass. The allowed regions explain the structure of the relationship between the information and disturbance including both the upper and lower bounds on the information for a given disturbance, even though the lower bounds can be violated by non-quantum effects such as classical noise and the observer’s non-optimal estimation. In particular, optimal measurements saturating the upper bounds are shown to be different for the four types of information–disturbance pairs. Therefore, our results broaden our understanding of the effects of quantum measurements and provide a useful tool for quantum information processing and communication.

Two of the above bounds have been shown by Banaszek \cite{7} and Cheong and Lee \cite{25} to be inequalities for the average values via different methods than ours. The most important difference is that they directly discussed the information and disturbance averaged over outcomes, whereas we start with those pertaining to each single outcome derived \cite{26} in the context of a physically reversible measurement \cite{27,29}. Even though trade-offs between information and disturbance are conventionally discussed using the average values \cite{6,7,9,10,16,18}, physically reversible measurements strongly imply trade-offs at the level of a single outcome \cite{11}. That is, in a physically
reversible measurement, whenever a second measurement called the reversing measurement recovers the pre-measurement state of the first measurement, it erases all the information obtained by the first measurement (see the Erratum of Ref. [35]). This state recovery with information erasure occurs not on average but only when the reversing measurement yields a preferred single outcome.

Moreover, starting from the level of a single outcome greatly simplifies the derivation of the allowed regions and optimal measurements. It is easy to show the allowed regions pertaining to a single outcome because the information and disturbance pertaining to a single outcome contain only a definite number of bounded parameters and have some useful invariances under parameter transformations. From these allowed regions, the allowed regions for the average values are shown using a graphical method based on an analogy with the center of mass, which makes it easy to construct the optimal measurements. In fact, without our method, it would be difficult to find all of the bounds and optimal measurements.

The rest of this paper is organized as follows. Section 2 reviews the procedure for quantifying the information and disturbances in quantum measurements. Sections 3 and 4 show the allowed regions for information and disturbance pertaining to a single outcome and those for the average values over all possible outcomes. Section 5 discusses the optimal measurements to show their differences for the four types of information–disturbance pairs. Section 6 summarizes our results.

2 Information and Disturbance

First, the amount of information provided by a measurement is quantified. Suppose that the \( d \)-level system to be measured is known to be in one of a set of predefined pure states \( \{ | \psi(a) \rangle \} \). The probability for \( | \psi(a) \rangle \) is given by \( p(a) \); however, which \( | \psi(a) \rangle \) is actually assigned to the system is unknown. Here we focus on the case where no prior information concerning the system is available, assuming that \( \{ | \psi(a) \rangle \} \) is a set of all the possible pure states and that \( p(a) \) is uniform according to the normalized invariant measure over the pure states. Because \( \{ | \psi(a) \rangle \} \) in this case is a continuous set of states, the index \( a \) actually represents a set of continuous parameters such as the hyperspherical coordinates in \( 2d \) dimensions as in Ref. [26], where the summation over \( a \) is replaced with an integral over the coordinates using the
hyperspherical volume element.

It is measured to obtain information about the state of the system. A quantum measurement can be described by a set of measurement operators \( \{ \hat{M}_m \} \) that satisfy
\[
\sum_m \hat{M}_m^\dagger \hat{M}_m = \hat{I},
\]
where \( m \) denotes the outcome of the measurement and \( \hat{I} \) is the identity operator. Here, the quantum measurement has been assumed to be ideal [36] or efficient [8] in the sense that it does not have classical noise yielding mixed post-measurement states because we focus on the quantum nature of the measurement. When the system is in a state \( |\psi(a)\rangle \), the measurement \( \{ \hat{M}_m \} \) yields an outcome \( m \) with probability
\[
p(m|a) = \langle \psi(a)| \hat{M}_m^\dagger \hat{M}_m |\psi(a)\rangle,
\]
changing the state into
\[
|\psi(m,a)\rangle = \frac{1}{\sqrt{p(m|a)}} \hat{M}_m |\psi(a)\rangle.
\]
Each measurement operator can be decomposed by a singular-value decomposition, such as
\[
\hat{M}_m = \hat{U}_m \hat{D}_m \hat{V}_m,
\]
where \( \hat{U}_m \) and \( \hat{V}_m \) are unitary operators and \( \hat{D}_m \) is a diagonal operator in an orthonormal basis \( \{|i\rangle\} \) with \( i = 1, 2, \ldots, d \) such that
\[
\hat{D}_m = \sum_i \lambda_{mi} |i\rangle \langle i|.
\]
The diagonal elements \( \{\lambda_{mi}\} \) are called the singular values of \( \hat{M}_m \) and satisfy \( 0 \leq \lambda_{mi} \leq 1 \).

From the outcome \( m \), the state of the system can be partially deduced. For example, Bayes’s rule states that, given an outcome \( m \), the probability that the state was \( |\psi(a)\rangle \) is given by
\[
p(a|m) = \frac{p(m|a) \ p(a)}{p(m)},
\]
where \( p(m) \) is the total probability of outcome \( m \),
\[
p(m) = \sum_a p(m|a) p(a).
\] (7)

That is, the outcome \( m \) changes the probability distribution for the states from \( \{p(a)\} \) to \( \{p(a|m)\} \). This change decreases the Shannon entropy, which is known as a measure of the lack of information:
\[
I(m) = \left[ -\sum_a p(a) \log_2 p(a) \right] - \left[ -\sum_a p(a|m) \log_2 p(a|m) \right].
\] (8)

Therefore, \( I(m) \), which we define as the information gain, quantifies the amount of information provided by the outcome \( m \) of the measurement \( \{\hat{M}_m\} \) and is explicitly written in terms of the singular values of \( \hat{M}_m \) as
\[
I(m) = \log_2 d - \frac{1}{\ln 2} \left[ \eta(d) - 1 \right] - \log_2 \sigma_m^2 + \frac{1}{\sigma_m^2} \sum_i \frac{\lambda_{mi}^2 \log_2 \lambda_{mi}^2}{\prod_{k \neq i} (\lambda_{mi}^2 - \lambda_{mk}^2)},
\] (9)

where
\[
\eta(n) = \sum_{k=1}^n \frac{1}{k}, \quad \sigma_m^2 = \sum_i \lambda_{mi}^2.
\] (10)

Note that \( I(m) \) satisfies
\[
0 \leq I(m) \leq \log_2 d - \frac{1}{\ln 2} [\eta(d) - 1].
\] (11)

The average of \( I(m) \) over all outcomes,
\[
I = \sum_m p(m) I(m),
\] (12)

is equal to the mutual information \( \[6\] \) between the random variables \( \{a\} \) and \( \{m\} \),
\[
I = \sum_{m,a} p(m,a) \log_2 \frac{p(m,a)}{p(m) p(a)}
\] (13)
with \( p(m, a) = p(m|a) p(a) \) because \( p(a) \) is uniform.

Alternatively, the state of the system can be estimated as a state \( |\varphi(m)\rangle \) depending on the outcome \( m \). In the optimal estimation \([7]\), \( |\varphi(m)\rangle \) is the eigenvector of \( \hat{M}_m^\dagger \hat{M}_m \) corresponding to its maximum eigenvalue. The quality of the estimate is evaluated by the estimation fidelity such that

\[
G(m) = \sum_a p(a|m) |\langle \varphi(m)|\psi(a)\rangle|^2. \tag{14}
\]

As was found for \( I(m) \), \( G(m) \) also quantifies the amount of information provided by the outcome \( m \) of the measurement \( \{\hat{M}_m\} \) [cf. Eq. (8)] and is explicitly written in terms of the singular values of \( \hat{M}_m \) as \([20]\)

\[
G(m) = \frac{1}{d + 1} \left( \frac{\sigma_m^2 + \lambda_{m,\text{max}}^2}{\sigma_m^2} \right), \tag{15}
\]

where \( \lambda_{m,\text{max}} \) is the maximum singular value of \( \hat{M}_m \). Note that \( G(m) \) satisfies

\[
\frac{1}{d} \leq G(m) \leq \frac{2}{d + 1}. \tag{16}
\]

The average of \( G(m) \) over all outcomes,

\[
G = \sum_m p(m) G(m), \tag{17}
\]

becomes the mean estimation fidelity discussed in Ref. \([7]\) because

\[
p(m) = \frac{\sigma_m^2}{d}, \quad \sum_m \sigma_m^2 = d, \tag{18}
\]

even though \( G(m) \) was not derived in Ref. \([7]\). Note that \( G \) can be derived from \( G(m) \); however, \( G(m) \) cannot be derived from \( G \). That is, \( G(m) \) characterizes the measurement \( \{\hat{M}_m\} \) in more detail than \( G \).

Next, the degree of disturbance caused by the measurement is quantified. When the measurement \( \{\hat{M}_m\} \) yields an outcome \( m \), the state of the system changes from \( |\psi(a)\rangle \) to \( |\psi(m, a)\rangle \), as given in Eq. (3). The size of this state change is evaluated by the operation fidelity such that

\[
F(m) = \sum_a p(a|m) |\langle \psi(a)|\psi(m, a)\rangle|^2. \tag{19}
\]
\[ F(m) \] quantifies the degree of disturbance caused when the measurement \( \{ \hat{M}_m \} \) yields the outcome \( m \) and is explicitly written in terms of the singular values of \( \hat{M}_m \) as \[ F(m) = \frac{1}{d+1} \left( \frac{\sigma_m^2 + \tau_m^2}{\sigma_m^2} \right), \] (20)

where
\[ \tau_m = \sum_i \lambda_{mi}, \] (21)

Note that \( F(m) \) satisfies
\[ \frac{2}{d+1} \leq F(m) \leq 1. \] (22)

Similar to \( G(m) \), the average of \( F(m) \) over all outcomes,
\[ F = \sum_m p(m) F(m), \] (23)

becomes the mean operation fidelity discussed in Ref. [7], even though \( F(m) \) was not derived in Ref. [7].

In addition to the size of the state change, the reversibility of the state change can also be regarded as a measure of the disturbance. Even though \( |\psi(a)\rangle \) and \( |\psi(m,a)\rangle \) are unknown, this state change is physically reversible if \( \hat{M}_m \) has a bounded left inverse \( \hat{M}_m^{-1} \) [28][29]. To recover \( |\psi(a)\rangle \), a second measurement called a reversing measurement is made on \( |\psi(m,a)\rangle \). The reversing measurement is described by another set of measurement operators \( \{ \hat{R}_\mu^{(m)} \} \) that satisfy
\[ \sum_{\mu} \hat{R}_{\mu}^{(m)\dagger} \hat{R}_{\mu}^{(m)} = \hat{I}, \] (24)

and, moreover, \( \hat{R}_{\mu_0}^{(m)} \propto \hat{M}_m^{-1} \) for a particular \( \mu = \mu_0 \), where \( \mu \) denotes the outcome of the reversing measurement. When the reversing measurement yields the preferred outcome \( \mu_0 \), the state of the system reverts to \( |\psi(a)\rangle \) via the state change caused by the reversing measurement because \( \hat{R}_{\mu_0}^{(m)} \hat{M}_m \propto \hat{I} \). For the optimal reversing measurement [34], the probability of recovery is given by
\[ R(m,a) = \frac{\lambda_{m,\text{min}}^2}{p(m|a)}, \] (25)
where $\lambda_{m,\min}$ is the minimum singular value of $\hat{M}_m$. The reversibility of the state change is then evaluated by this maximum successful probability as

$$R(m) = \sum_a p(a|m) R(m, a). \quad (26)$$

As was found for $F(m)$, $R(m)$ also quantifies the degree of disturbance caused when the measurement $\{\hat{M}_m\}$ yields the outcome $m$ [cf. Eq. (19)] and is explicitly written in terms of the singular values of $\hat{M}_m$ as

$$R(m) = d \left( \frac{\lambda_{m,\min}^2}{\sigma_m^2} \right). \quad (27)$$

Note that $R(m)$ satisfies

$$0 \leq R(m) \leq 1. \quad (28)$$

The average of $R(m)$ over all outcomes,

$$R = \sum_m p(m) R(m), \quad (29)$$

is the degree of physical reversibility of a measurement discussed in Ref. [34], whose explicit form in terms of the singular values is given in Ref. [25], even though $R(m)$ was not derived in Ref. [25].

Therefore, the information and disturbance for a single outcome $m$ are obtained as functions of the singular values of $\hat{M}_m$: $I(m)$ and $G(m)$ for the information and $F(m)$ and $R(m)$ for the disturbance. Note that they are invariant under the interchange of any pair of singular values,

$$\lambda_{mi} \longleftrightarrow \lambda_{mj} \quad \text{for any} \ (i, j), \quad (30)$$

and under rescaling of all the singular values,

$$\lambda_{mi} \longrightarrow c\lambda_{mi} \quad \text{for all} \ i, \quad (31)$$

by a constant $c$ [26]. By contrast, the probability for the outcome $m$, $p(m) = \sigma_m^2/d$, is invariant under the interchange but is not invariant under the rescaling.

As an important example, consider $\hat{M}_{k,l}^{(d)}(\lambda)$, which is defined as a measurement operator whose singular values are

$$\underbrace{1, 1, \ldots, 1}_{k}, \lambda, \underbrace{\lambda, \ldots, \lambda}_{l}, \underbrace{0, 0, \ldots, 0}_{d-k-l} \quad (32)$$
with $0 \leq \lambda \leq 1$. Even though the information and disturbance for $\hat{M}_{k,l}(0)$ can be calculated from Eqs. (9), (15), (20), and (27), calculating $I(m)$ is not straightforward due to the degeneracy of the singular values. By taking the limit $\lambda_{mi} \rightarrow \lambda_{mk}$, $I(m)$ is found to be [26]

$$I(m) = \log_2 d - \frac{1}{\ln 2} \left[ \eta(d) - 1 \right] - \log_2 (k + \lambda^2)$$

$$+ \frac{1}{k + \lambda^2} \left[ \lambda^{2(k+1)} \log_2 \lambda^2 - \sum_{n=0}^{k-1} \frac{a_n^{(k+1)}}{(\lambda^2 - 1)^{k-n}} \right]$$

(33)

for $\hat{M}_{k,l}(\lambda)$ and

$$I(m) = \log_2 d - \frac{1}{\ln 2} \left[ \eta(d) - 1 \right]$$

$$- \log_2 (1 + l\lambda^2) - \frac{1}{1 + l\lambda^2} \sum_{n=0}^{l-1} \frac{c_n^{(l+1)}(\lambda)}{(1 - \lambda^2)^l - n}$$

(34)

for $\hat{M}_{l,l}(\lambda)$, where $\{a_n^{(j)}\}$ and $\{c_n^{(j)}(\lambda)\}$ are given by

$$a_n^{(j)} = \left[ \frac{1}{\ln 2} \right] \binom{j}{n} \left[ \eta(j) - \eta(j - n) \right],$$

(35)

$$c_n^{(j)}(\lambda) = \lambda^{2(j-n)} \left[ \binom{j}{n} \log_2 \lambda^2 + a_n^{(j)} \right].$$

(36)

Similarly, $\hat{P}_r^{(d)}$ is defined as a projective measurement operator of rank $r$. Note that $\hat{M}_{k,l}(0) = \hat{P}_k$, $\hat{M}_{k,l}(1) = \hat{P}_{k+l}$, and $\hat{P}_r^{(d)} = \hat{I}$. For $\hat{P}_r^{(d)}$, $I(m)$ is found to be [38]

$$I(m) = \log_2 \frac{d}{r} - \frac{1}{\ln 2} \left[ \eta(d) - \eta(r) \right].$$

(37)

### 3 Allowed Region

Next, we plot the information and disturbance for various measurement operators on a plane. In particular, an allowed region for information versus disturbance can be shown on the plane by plotting all physically possible measurement operators; that is, by varying every singular value over the range of $0 \leq \lambda_{mi} \leq 1$. It is easy to do this for $I(m)$, $G(m)$, $F(m)$, and
$R(m)$ because they contain only a definite number of bounded parameters, i.e., $d$ singular values, in contrast to $I$, $G$, $F$, and $R$. Moreover, from the interchange invariance in Eq. (30), measurement operators having the same singular values up to ordering correspond to the same point on the plane. According to the rescaling invariance in Eq. (31), $\hat{M}_m$ and $c\hat{M}_m$ correspond to the same point on the plane.

Figure 1(a) shows the allowed region for $G(m)$ versus $F(m)$ when $d = 4$ in blue (dark gray). In the figure, $P_r$ and $(k,l)$ represent the point corresponding to $c\hat{P}_r^{(d)}$ and the line corresponding to $c\hat{M}_{k,1}^{(d)}(\lambda)$ with $0 \leq \lambda \leq 1$, respectively. The upper boundary consists of one curved line $(1,d-1)$ connecting $P_1$ and $P_d$ as $\lambda$ varies from 0 to 1, whereas the lower boundary consists of $d - 1$ curved lines $(k,1)$ connecting $P_k$ to $P_{k+1}$ for $k = 1,2,\ldots,d-1$. Conversely, Fig. 1(b) shows the allowed region for $G(m)$ versus $R(m)$ when $d = 4$ in blue (dark gray). In this case, both the upper and lower boundaries consist of one straight line: $(1,d-1)$ for the upper boundary and $(d-1,1)$ for the lower boundary. Similarly, Figs. 1(c) and 1(d) show the allowed region for $I(m)$ versus $F(m)$ and for $I(m)$ versus $R(m)$, respectively. The measurement operators corresponding to the upper and lower boundaries are the same as for $G(m)$, even though the lines have different shapes. Figure 2 shows the allowed regions when $d = 8$ in blue (dark gray).

The above boundaries, $(1,d-1)$ and $(k,1)$, were first confirmed by brute-force numerical calculations where every singular value was varied by steps of $\Delta \lambda_{mi} = 0.01$ for $d = 2,3,\ldots,6$ and $\Delta \lambda_{mi} = 0.02$ for $d = 7,8$. Moreover, for $G(m)$ versus $F(m)$ and for $G(m)$ versus $R(m)$, the boundaries can analytically be proven to be the true boundaries for arbitrary $d$ (see Appendix A). Unfortunately, however, for $I(m)$ versus $F(m)$ and for $I(m)$ versus $R(m)$, proving that the boundaries are the true boundaries is difficult analytically. Nevertheless, they can be shown to satisfy the necessary conditions for the true boundaries using the Karush–Kuhn–Tucker (KKT) conditions [39], which generalize the method of Lagrange multipliers to handle inequality constraints in mathematical optimization. For example, to find the lower boundary for $I(m)$ versus $F(m)$, consider minimizing $I(m)$ subject to $F(m) = F_0$ and $\lambda_{mi} \geq 0$ ($i = 1,2,\ldots,d$). Then, $\hat{M}_{k,1}^{(d)}(\lambda)$ satisfies a necessary condition for a local minimum, that is, for a Lagrange function

$$L_F = I(m) - \alpha_F [F(m) - F_0] - \sum_i \beta_i \lambda_{mi},$$

(38)

$\hat{M}_{k,1}^{(d)}(\lambda)$ satisfies $\partial L_F / \partial \lambda_{mi} = 0$ with KKT multipliers $\alpha_F$ and $\{\beta_i\}$ such that
Figure 1: Four allowed regions for information versus disturbance for $d = 4$: (a) estimation fidelity $G(m)$ versus operation fidelity $F(m)$, (b) estimation fidelity $G(m)$ versus physical reversibility $R(m)$, (c) information gain $I(m)$ versus operation fidelity $F(m)$, and (d) information gain $I(m)$ versus physical reversibility $R(m)$. In each panel, the region pertaining to a single outcome is shown in blue (dark gray), and the extended region obtained by averaging over all outcomes is shown in yellow (light gray).
Figure 2: Four allowed regions for information versus disturbance for $d = 8$: (a) estimation fidelity $G(m)$ versus operation fidelity $F(m)$, (b) estimation fidelity $G(m)$ versus physical reversibility $R(m)$, (c) information gain $I(m)$ versus operation fidelity $F(m)$, and (d) information gain $I(m)$ versus physical reversibility $R(m)$. In each panel, the region pertaining to a single outcome is shown in blue (dark gray), and the extended region obtained by averaging over all outcomes is shown in yellow (light gray).
\[ \beta_i \geq 0 \text{ and } \beta_i \lambda_{mi} = 0 \text{ for all } i \] and has \( \lambda = \lambda_0 \) such that \( F(m) = F_0 \) if \( (k + 1)/(d + 1) \leq F_0 \leq (k + 2)/(d + 1) \). These mathematical optimizations are explained in Appendix B.

## 4 Average over Outcomes

Here, the regions that are allowed for the information and disturbance averaged over all possible outcomes are discussed: \( I \) and \( G \) for the information and \( F \) and \( R \) for the disturbance. Unfortunately, it is difficult to show the allowed regions directly from their explicit forms written in terms of the singular values because the number of singular values contained in them is not definite due to the indefinite number of outcomes. Note that there are no physical limitations on the number of outcomes.

Instead, we show the allowed regions using the following analogy with the center of mass. In the measurement \( \{ \hat{M}_m \} \), each measurement operator \( \hat{M}_m \) corresponds to a point \( R_m \) in the allowed region pertaining to a single outcome with weight \( p(m) \). This situation can be viewed as a set of particles, each with a mass \( p(m) \) located at a point \( R_m \). The center of mass of these particles then indicates the average information and disturbance of the measurement. Conversely, for an arbitrary set of particles located in the allowed region pertaining to a single outcome, an equivalent measurement satisfying Eq. (1) can be constructed by rescaling and duplicating the measurement operators, as shown in Appendix C. For example, for \( d = 4 \), two particles with the same mass \( 1/2 \) located at \( P_1 \) and \( P_4 \) in Fig. 1 can be simulated by a measurement with five outcomes whose measurement operators are

\[
\hat{M}_m = \begin{cases} 
\frac{1}{\sqrt{2}} |m\rangle\langle m| & (m = 1, 2, 3, 4) \\
\frac{1}{\sqrt{2}} \hat{I} & (m = 5).
\end{cases}
\]

Therefore, the allowed region for the average information and disturbance can be shown by considering the center of mass of all possible sets of particles. Note that the center of mass may be located outside the region where the particles are situated, which means that the allowed region is extended by averaging over the outcomes. The resultant region is the convex hull of the original region.

The regions extended by averaging are shown in Fig. 1 in yellow (light gray). As shown in Fig. (a), the lower boundary for \( G \) versus \( F \) is extended
Figure 3: Two line slopes for $d = 4$. $D_4(\lambda)$ is the slope of the tangent line to the line $(1, 3)$ at a point $Q$, and $S_4(\lambda)$ is the slope of the straight line from $P_4$ to $Q$. Note that the horizontal axis is reversed.

to the straight lines between $P_k$ and $P_{k+1}$ for $k = 1, 2, \ldots, d - 1$, whereas the upper boundary is not extended due to its convexity. By contrast, as shown in Fig. 1(b), the boundaries for $G$ versus $R$ are not extended at all. Meanwhile, as shown in Fig. 1(c), the lower boundary for $I$ versus $F$ is extended as in the case of $G$ and, moreover, the upper boundary is extended a little higher when $d \geq 3$ because the line $(1, d - 1)$ has a slight dent near $P_d$. In fact, an analytic calculation of $\hat{M}_{1,d-1}(\lambda)$ shows that

$$\frac{d^2 F(m)}{dI(m)^2} > 0$$

near $P_d$ when $d \geq 3$. The upper boundary is therefore extended to the tangent line drawn from $P_d$ to the line $(1, d - 1)$ between $P_d$ and the point of tangency $T$. As shown in Fig. 1(d), the upper boundary for $I$ versus $F$ is extended to the straight line between $P_1$ and $P_d$, whereas the lower boundary is not extended. The case of $d = 8$ is shown in Fig. 2.

To find the point $T$ on the upper boundary for $I$ versus $F$, two line slopes are defined as functions of $\lambda$: the slope of the tangent line to the line $(1, d - 1)$ at the point $Q$ corresponding to $\hat{M}_{1,d-1}^{(d)}(\lambda)$,

$$D_d(\lambda) = \frac{dF(m)}{dI(m)},$$

(41)
Figure 4: Singular value $\lambda_T$, information $I_T$, and fidelity $F_T$ at the point of tangency $T$ for various $d$.

and the slope of the straight line from $P_d$ to $Q$,

$$S_d(\lambda) = \frac{F(m) - 1}{I(m)}.$$  \hspace{1cm} (42)

These functions are shown for $d = 4$ in Fig. 3. Using $\lambda_T$ such that

$$D_d(\lambda_T) = S_d(\lambda_T),$$  \hspace{1cm} (43)

the measurement operator corresponding to $T$ can be written as $\hat{M}_{1,d-1}^{(d)}(\lambda_T)$. In Fig. 4 $\lambda_T$ is shown with $I(m)$ and $F(m)$ at $T$, denoted by $I_T$ and $F_T$, respectively, for various $d$. When $d = 4$, $T$ in Fig. 4(c) corresponds to $\hat{M}_{1,3}^{(4)}(0.299)$ and the upper boundary for $I$ versus $F$ moves up between $P_4$ and $T$, at most by $3.5 \times 10^{-3}$. This extension of the upper boundary becomes larger as $d$ increases. For example, when $d = 8$, $T$ in Fig. 2(c) corresponds to $\hat{M}_{1,7}^{(8)}(0.120)$ and the upper boundary moves up at most by $2.6 \times 10^{-2}$.

Interestingly, $\hat{M}_{1,d-1}^{(d)}(\lambda_T)$ is the most efficient measurement operator in terms of the ratio of information gain to fidelity loss [26],

$$E_F(m) = \frac{I(m)}{1 - F(m)}.$$  \hspace{1cm} (44)

The upper boundary for $G$ versus $F$ and that for $G$ versus $R$ are equivalent to the inequalities of Banaszek [7] and Cheong and Lee [25], respectively,
where the averages are explicitly calculated using $p(m) = \sigma^2_m/d$. However, to our knowledge, this is the first derivation of the other two upper and four lower boundaries. The lower boundaries are less important than the upper boundaries in quantum information and can be violated by non-ideal measurements, which have classical noise yielding mixed post-measurement states, or by non-optimal estimations, which assume unwise observers making incorrect choices for $|\varphi(m)\rangle$ in $G(m)$. Nevertheless, for the foundations of quantum mechanics, it is worth deriving both the upper and lower boundaries for ideal measurements with optimal estimation to examine the intrinsic nature and power of quantum measurements.

The case of $d = 2$ is a special case, where the regions extended by averaging are the main parts of the allowed regions, as shown in Fig. 5. In this case, the allowed regions pertaining to a single outcome shrink to the line $(1,1)$ because a measurement operator can be represented by a single parameter via the rescaling invariance in Eq. (31) [24]. Moreover, the line $(1,1)$ in Fig. 5(c) has no dent unlike the case of $d \geq 3$. In fact, it can be shown for $\hat{M}_{1,1}^{(2)}(\lambda)$ that

$$\frac{d^2 F(m)}{dI(m)^2} < 0 \quad (45)$$

near $P_2$. The point T does not exist on the line $(1,1)$ because the slopes $D_2(\lambda)$ and $S_2(\lambda)$ in Eqs. (41) and (42) do not become equal to each other except for $\lambda = 1$, as shown in Fig. 6.

5 Optimal Measurement

Finally, we discuss the optimal measurements saturating the upper bounds on the information for a given disturbance. The upper bounds are denoted by the upper boundaries of the allowed regions for the average information and disturbance. Therefore, according to the analogy with the center of mass, a measurement is optimal for an information–disturbance pair if it is equivalent to a set of particles whose center of mass is on the upper boundary for that information–disturbance pair. The optimal measurements are different for the four types of information–disturbance pairs because the upper boundaries have different shapes on the four information–disturbance planes, as shown in Fig. 1.

The conditions for the optimal measurements are as follows. A measurement $\{\hat{M}_m\}$ is optimal for $G$ versus $F$ if all $\hat{M}_m$’s correspond to an identical
Figure 5: Four allowed regions for information versus disturbance for $d = 2$: (a) estimation fidelity $G(m)$ versus operation fidelity $F(m)$, (b) estimation fidelity $G(m)$ versus physical reversibility $R(m)$, (c) information gain $I(m)$ versus operation fidelity $F(m)$, and (d) information gain $I(m)$ versus physical reversibility $R(m)$. In each panel, the region pertaining to a single outcome is just the solid line denoted by $(1,1)$ and the extended region obtained by averaging over all outcomes is shown in yellow (light gray).
Figure 6: Two line slopes for $d = 2$. $D_2(\lambda)$ is the slope of the tangent line to the line $(1,1)$ at a point $Q$, and $S_2(\lambda)$ is the slope of the straight line from $P_2$ to $Q$. Note that the horizontal axis is reversed.
Figure 7: Four conditions for optimal measurements. For example, the set $G$-$F$ represents all measurements that are optimal for $G$ versus $F$.

...ments satisfy all four conditions.

As a specific example, consider a measurement \( \{\hat{M}_m^{(d)}(\lambda)\} \) with \( d \) outcomes, \( m = 1, 2, \ldots, d \), where \( \hat{M}_m^{(d)}(\lambda) \) is defined by

\[
\hat{M}_m^{(d)}(\lambda) = \frac{1}{\sqrt{1 + (d - 1)\lambda^2}} \left( |m\rangle \langle m| + \sum_{i \neq m} \lambda|i\rangle \langle i| \right) \tag{46}
\]

with \( 0 < \lambda < 1 \). For a given \( \lambda \), all \( \hat{M}_m^{(d)}(\lambda) \)'s correspond to an identical point on the line \((1, d-1)\) in the four information–disturbance planes because they are equivalent to \( \hat{M}_{1,d-1}^{(d)}(\lambda) \) via the interchange and rescaling invariances in Eqs. (30) and (31). The corresponding point on the line \((1, d-1)\) indicates the average information and disturbance of \( \{\hat{M}_m^{(d)}(\lambda)\} \). The measurement \( \{\hat{M}_m^{(d)}(\lambda)\} \) is optimal both for $G$ versus $F$ and for $G$ versus $R$ for arbitrary \( \lambda \) because the line \((1, d-1)\) is equal to the upper boundary, as shown in Figs. II(a) and II(b).

However, the measurement \( \{\hat{M}_m^{(d)}(\lambda)\} \) is not necessarily optimal for $I$ versus $F$ because only a part of the line \((1, d-1)\) is equal to the upper boundary when \( d \geq 3 \), as shown in Fig. II(c). It is optimal for $I$ versus $F$ only if \( \lambda \leq \lambda_T \), with \( \lambda_T \) being defined by Eq. (43). Note that \( \hat{M}_m^{(d)}(\lambda_T) \) corresponds to \( T \) on the S-shaped curve \((1, d-1)\). If \( \lambda > \lambda_T \), \( \hat{M}_m^{(d)}(\lambda) \) corresponds to a point on the concave part between \( P_d \) and \( T \) of the line \((1, d-1)\), where the upper boundary is equal to the straight line between \( P_d \) and \( T \). This means that \( \{\hat{M}_m^{(d)}(\lambda)\} \) is not optimal for $I$ versus $F$ if \( \lambda > \lambda_T \) or equivalently if \( F > F_T \). The optimal measurement for this case can easily
be constructed from the analogy with the center of mass by considering two particles: one located at $T$ with mass $q$ and the other located at $P_d$ with mass $1 - q$. According to Appendix C, the optimal measurement has $d + 1$ outcomes whose measurement operators are

$$
\hat{M}_m = \begin{cases} 
\sqrt{q} \hat{M}_m^{(d)}(\lambda_T) & (m = 1, 2, \ldots, d) \\
\sqrt{1 - q} \hat{I} & (m = d + 1),
\end{cases}
$$

(47)

where $q = (1 - F) / (1 - F_T)$ for a given $F$. The average information and disturbance of this measurement are then indicated by a point on the straight line between $P_d$ and $T$ equal to a part of the upper boundary. By contrast, when $d = 2$, $\{\hat{M}_m^{(2)}(\lambda)\}$ is optimal for $I$ versus $F$ for arbitrary $\lambda$ because the line $(1, 1)$ is equal to the upper boundary, as shown in Fig. 5(c).

Conversely, the measurement $\{\hat{M}_m^{(d)}(\lambda)\}$ is not optimal for $I$ versus $R$ for any $\lambda$ because the line $(1, d - 1)$ is not equal to the upper boundary at all, as shown in Fig. 1(d). In this case, the upper boundary is the straight line between $P_d$ and $P_1$. Therefore, the optimal measurement for $I$ versus $R$ can be constructed from the analogy with the center of mass by considering two particles: one located at $P_1$ with mass $q$ and the other located at $P_d$ with mass $1 - q$. This has $d + 1$ outcomes whose measurement operators are

$$
\hat{M}_m = \begin{cases} 
\sqrt{q} |m\rangle\langle m| & (m = 1, 2, \ldots, d) \\
\sqrt{1 - q} \hat{I} & (m = d + 1),
\end{cases}
$$

(48)

where $q = 1 - R$ for a given $R$. The average information and disturbance of this measurement are indicated by a point on the straight line between $P_d$ and $P_1$ equal to the upper boundary.

Of course, the measurements given in Eqs. (47) and (48) are also optimal for $G$ versus $R$ for arbitrary $q$. Even though their measurement operators correspond to different points on the line $(1, d - 1)$, the point indicating the average values is still on the line $(1, d - 1)$ equal to the upper boundary because the line $(1, d - 1)$ is straight, as shown in Fig. 1(b). However, except for $q = 0$ or 1, the measurement in Eq. (47) is optimal neither for $G$ versus $F$ nor for $I$ versus $R$ and the measurement in Eq. (48) is optimal neither for $G$ versus $F$ nor for $I$ versus $F$. 

20
6 Summary

In summary, we have shown the allowed regions for information versus disturbance for quantum measurements of completely unknown states. The information and disturbances pertaining to a single outcome are quantified using the singular values of the measurement operator and are plotted on four types of information–disturbance planes to show the allowed regions pertaining to a single outcome. The allowed regions for the average values are also discussed via an analogy with the center of mass. These regions explicitly give not only the upper bounds but also the lower bounds on the information for a given disturbance together with the optimal measurements saturating the upper bounds. Consequently, our results broaden our perspective of quantum measurements and provide a useful tool for quantum information processing and communication.

Appendix

A Proof of Boundaries

Here, the proofs of the boundaries are outlined for $G(m)$ versus $F(m)$ and for $G(m)$ versus $R(m)$. To prove the upper and lower boundaries for $G(m)$ versus $F(m)$, consider maximizing and minimizing $F(m)$ for a given $G(m)$. Using the interchange and rescaling invariances in Eqs. (30) and (31), the singular values are assumed to be sorted in descending order, $\lambda_{m1} \geq \lambda_{m2} \geq \cdots \geq \lambda_{md}$, and normalized such that $\sigma_{m}^2 = 1$. Then, the problems are simplified to maximizing and minimizing $\sum_{i=2}^{d} \lambda_{mi}$ subject to $\sum_{i=2}^{d} \lambda_{mi}^2 = 1 - \lambda_{m1}^2$ and $0 \leq \lambda_{mi} \leq \lambda_{m1}$ for a given $\lambda_{m1}$ from Eqs. (15) and (20). The maximum is achieved when $\lambda_{m2} = \lambda_{m3} = \cdots = \lambda_{md} = \sqrt{(1-\lambda_{m1}^2)/(d-1)}$. The corresponding singular values are proportional to those of $\hat{M}_{1,d-1}^{(d)}(\lambda)$ with $\lambda = \sqrt{(1-\lambda_{m1}^2)/(d-1)}/\lambda_{m1}$. Therefore, the line $(1,d-1)$ is the upper boundary for $G(m)$ versus $F(m)$.

Conversely, the minimum is achieved when $\lambda_{m2} = \sqrt{1-\lambda_{m1}^2}$ and the others are 0 if $\lambda_{m1} \geq 1/\sqrt{2}$. Because these singular values are proportional to those of $\hat{M}_{1,1}^{(d)}(\lambda)$ with $\lambda = \sqrt{1-\lambda_{m1}^2}/\lambda_{m1}$, the line $(1,1)$ is the lower boundary for $G(m)$ versus $F(m)$ if $G(m) \geq 3/(2d + 2)$. However, if $\lambda_{m1} < 1/\sqrt{2}$, they do not satisfy $\lambda_{m2} \leq \lambda_{m1}$ because $\sqrt{1-\lambda_{m1}^2} > \lambda_{m1}$. Therefore,
in this case, let \( \lambda_{m2} = \lambda_{m1} \) and consider minimizing \( \sum_{i=3}^{d} \lambda_{mi} \) subject to \( \sum_{i=3}^{d} \lambda_{mi}^2 = 1 - 2\lambda_{m1}^2 \) and \( 0 \leq \lambda_{mi} \leq \lambda_{m1} \) for a given \( \lambda_{m1} \). If \( \lambda_{m1} \geq 1/\sqrt{3} \), the minimum is achieved when \( \lambda_{m3} = \sqrt{1 - 2\lambda_{m1}^2} \) and the others are 0. Because these singular values are proportional to those of \( \hat{M}_2^{(d)}(\lambda) \) with \( \lambda = \sqrt{1 - 2\lambda_{m1}^2}/\lambda_{m1} \), the line \((2, 1)\) is the lower boundary for \( G(m) \) versus \( F(m) \) if \( 4/(3d + 3) \leq G(m) < 3/(2d + 2) \). By repeating similar minimizations for \( \lambda_{m1} < 1/\sqrt{3} \), the lines \((k, 1)\) with \( k = 1, 2, \ldots, d - 1 \) are shown to be the lower boundaries for \( G(m) \) versus \( F(m) \).

Similarly, to prove the upper and lower boundaries for \( G(m) \) versus \( R(m) \), consider maximizing and minimizing \( R(m) \) for a given \( G(m) \). Via the descending ordering and the normalization \( \sigma_m^2 = 1 \), the problems are simplified to maximizing and minimizing \( \lambda_{md} \) subject to \( \sum_{i=2}^{d} \lambda_{mi}^2 = 1 - \lambda_{m1}^2 \) and \( 0 \leq \lambda_{mi} \leq \lambda_{m1} \) for a given \( \lambda_{m1} \) from Eqs. (15) and (27). As in the case of \( F(m) \), the maximum is achieved when \( \lambda_{m2} = \lambda_{m3} = \cdots = \lambda_{md} \). This result shows that the line \((1, d - 1)\) is the upper boundary for \( G(m) \) versus \( R(m) \). Conversely, the minimum is achieved when \( \lambda_{md} = 0 \) if \( \lambda_{m1} \geq 1/\sqrt{d - 1} \). That is, \( R(m) = 0 \) is the lower boundary for \( G(m) \) versus \( R(m) \) if \( G(m) \geq d/(d^2 - 1) \). However, if \( \lambda_{m1} < 1/\sqrt{d - 1} \), \( \lambda_{md} \) cannot be 0 to satisfy \( \sigma_m^2 = 1 \) because \( \lambda_{m2} \leq \lambda_{m1} \). In this case, the minimum is achieved when \( \lambda_{md} = \sqrt{1 - (d - 1)\lambda_{m1}^2} \) and the others are \( \lambda_{m1} \). These singular values are proportional to those of \( M_{d-1}^{(d)}(\lambda) \) with \( \lambda = \sqrt{1 - (d - 1)\lambda_{m1}^2}/\lambda_{m1} \). This result shows that the line \((d - 1, 1)\) is the lower boundary for \( G(m) \) versus \( R(m) \) if \( G(m) < d/(d^2 - 1) \).

**B Mathematical Optimization**

Here, the mathematical optimizations of the information for a given disturbance are outlined for \( I(m) \) versus \( F(m) \) and for \( I(m) \) versus \( R(m) \) based on the method of Lagrange multipliers and its generalization known as the Karush–Kuhn–Tucker (KKT) conditions \[39\]. Consider maximizing \( I(m) \) subject to \( F(m) = F_0 \) using a Lagrange function \( L^F = -I(m) - \alpha^F[F(m) - F_0] \) with a multiplier \( \alpha^F \). To use the method of Lagrange multipliers, the derivatives of \( I(m) \) and \( F(m) \) with respect to \( \lambda_{mi} \) should be calculated. From the rescaling invariance in Eq. (31), the derivatives of \( I(m) \) satisfy

\[
\sum_{i} \lambda_{mi} \frac{\partial I(m)}{\partial \lambda_{mi}} = 0
\]

(49)
according to Euler’s homogeneous function theorem. Using this equation and the interchange invariance in Eq. (50), the derivatives of $I(m)$ for $\dot{M}_{k,l}^{(d)}(\lambda)$ can be written as

$$
\frac{\partial I(m)}{\partial \lambda_{mi}} = \begin{cases} 
I_{k,l}^{(d)}(\lambda) & (1 \leq i \leq k) \\
-\frac{k}{l}I_{k,l}^{(d)}(\lambda) & (k + 1 \leq i \leq k + l) \\
0 & (k + l + 1 \leq i \leq d)
\end{cases}
$$

(50)

with $I_{k,l}^{(d)}(\lambda) \geq 0$, where the third case is 0 because $I(m)$ is a function of $\{\lambda_{mi}^2\}$. Similarly, the derivatives of $F(m)$ for $\dot{M}_{k,l}^{(d)}(\lambda)$ can be written as

$$
\frac{\partial F(m)}{\partial \lambda_{mi}} = \begin{cases} 
F_{k,l}^{(d)}(\lambda) & (1 \leq i \leq k) \\
-\frac{k}{l}F_{k,l}^{(d)}(\lambda) & (k + 1 \leq i \leq k + l) \\
\tilde{F}_{k,l}^{(d)}(\lambda) & (k + l + 1 \leq i \leq d)
\end{cases}
$$

(51)

with $F_{k,l}^{(d)}(\lambda) \leq 0$ and $\tilde{F}_{k,l}^{(d)}(\lambda) > 0$. These derivatives show that $\dot{M}_{i,d-1}^{(d)}(\lambda)$ satisfies $\partial F/\partial \lambda_{mi} = 0$ for all $i$ with a multiplier of $\alpha_F = -I_{1,d-1}^{(d)}(\lambda)/F_{1,d-1}^{(d)}(\lambda)$. Moreover, there exists a parameter $\lambda_0$ such that $F(m)$ for $\dot{M}_{1,d-1}^{(d)}(\lambda_0)$ is equal to $F_0$. That is, $\dot{M}_{1,d-1}^{(d)}(\lambda_0)$ satisfies a necessary condition for a local maximum according to the method of Lagrange multipliers. This result implies that the line $(1, d - 1)$ is the upper boundary for $I(m)$ versus $F(m)$.

Conversely, consider minimizing $I(m)$ subject to $F(m) = F_0$ and $\lambda_{mi} \geq 0$ ($i = 1, 2, \ldots, d$). The inequality constraints are indispensable in this case because the solutions are on the boundary of the parameter space, $\lambda_{mi} = 0$. To handle these inequality constraints, the KKT conditions are applied using a Lagrange function $L_F = I(m) - \alpha_F [F(m) - F_0] - \sum_i \beta_i \lambda_{mi}$ with multipliers $\alpha_F$ and $\{\beta_i\}$. Then, $\dot{M}_{i,1}^{(d)}(\lambda)$ satisfies $\partial L_F/\partial \lambda_{mi} = 0$ for all $i$ with multipliers $\alpha_F = I_{i,1}^{(d)}(\lambda)/F_{i,1}^{(d)}(\lambda)$ and

$$
\beta_i = \begin{cases} 
0 & (1 \leq i \leq k + 1) \\
-\alpha_F \tilde{F}_{i,1}^{(d)}(\lambda) & (k + 2 \leq i \leq d).
\end{cases}
$$

(52)

In addition, these $\{\beta_i\}$ satisfy the requirements as multipliers for the inequality constraints, $\beta_i \geq 0$ and $\beta_i \lambda_{mi} = 0$, for all $i$. There exists a parameter
\( \lambda_0 \) such that \( F(m) \) for \( \hat{M}_{k,1}^{(d)}(\lambda_0) \) is equal to \( F_0 \) if \( (k + 1)/(d + 1) \leq F_0 \leq (k + 2)/(d + 1) \). That is, \( \hat{M}_{k,1}^{(d)}(\lambda_0) \) satisfies a necessary condition for a local minimum according to the KKT conditions. This result implies that the line \((k,1)\) is the lower boundary for \( I(m) \) versus \( F(m) \).

Similarly, letting \( \lambda_{m,\min} = \lambda_{md} \), consider maximizing \( I(m) \) subject to \( R(m) = R_0 \) and \( \lambda_{mi} - \lambda_{md} \geq 0 \) \( (i = 1, 2, \ldots, d - 1) \) using a Lagrange function \( L^R = -I(m) - \alpha^R[R(m) - R_0] - \sum_i \gamma_i(\lambda_{mi} - \lambda_{md}) \) with multipliers \( \alpha^R \) and \( \{\gamma_i\} \). The derivatives of \( R(m) \) for \( \hat{M}_{k,l}^{(d)}(\lambda) \) can be written when \( k + l = d \) such that

\[
\frac{\partial R(m)}{\partial \lambda_{mi}} \equiv \begin{cases} \mathcal{R}_{k,l}^{(d)}(\lambda) & (1 \leq i \leq k) \\ -\frac{k}{l} \mathcal{R}_{k,l}^{(d)}(\lambda) - \frac{1-\delta_{k,l}}{l} \mathcal{R}_{k,l}^{\tilde{d}}(\lambda) & (k + 1 \leq i \leq d) \end{cases}
\]  

(53)

with \( \mathcal{R}_{k,l}^{(d)}(\lambda) \leq 0 \) and \( \mathcal{R}_{k,l}^{\tilde{d}}(\lambda) \geq 0 \). Then, \( \hat{M}_{1,d-1}^{(d)}(\lambda) \) satisfies \( \partial L^R / \partial \lambda_{mi} = 0 \) for all \( i \) with multipliers \( \alpha^R = -\mathcal{I}_{1,d-1}^{(d)}(\lambda)/\mathcal{R}_{1,d-1}^{(d)}(\lambda) \) and

\[
\gamma_i = \begin{cases} 0 & (i = 1) \\ \frac{1}{d-1} \alpha^R \mathcal{R}_{1,d-1}^{\tilde{d}}(\lambda) & (2 \leq i \leq d - 1) \end{cases}
\]  

(54)

satisfying \( \gamma_i \geq 0 \) and \( \gamma_i(\lambda_{mi} - \lambda_{md}) = 0 \) for all \( i \). Moreover, there exists a parameter \( \lambda_0 \) such that \( R(m) \) for \( \hat{M}_{1,d-1}^{(d)}(\lambda_0) \) is equal to \( R_0 \). According to the KKT conditions, \( \hat{M}_{1,d-1}^{(d)}(\lambda_0) \) satisfies a necessary condition for a local maximum implying that the line \((1,d-1)\) is the upper boundary for \( I(m) \) versus \( R(m) \). Conversely, consider minimizing \( I(m) \) subject to \( R(m) = R_0 \) using a Lagrange function \( L_R = I(m) - \alpha_R[R(m) - R_0] \) with a multiplier \( \alpha_R \). Then, \( \hat{M}_{d-1,1}^{(d)}(\lambda) \) satisfies \( \partial L_R / \partial \lambda_{mi} = 0 \) for all \( i \) with the multiplier \( \alpha_R = \mathcal{I}_{d-1,1}^{(d)}(\lambda)/\mathcal{R}_{d-1,1}^{(d)}(\lambda) \) and there exists a parameter \( \lambda_0 \) such that \( R(m) \) for \( \hat{M}_{d-1,1}^{(d)}(\lambda_0) \) is equal to \( R_0 \). According to the method of Lagrange multipliers, \( \hat{M}_{d-1,1}^{(d)}(\lambda_0) \) satisfies a necessary condition for a local minimum implying that the line \((d-1,1)\) is the lower boundary for \( I(m) \) versus \( R(m) \).

**C Construction of Equivalent Measurement**

Here, the general construction of an equivalent measurement is presented for an arbitrary set of particles located in the allowed region pertaining to
a single outcome. The construction is not trivial because a measurement operator not only corresponds to a point but also gives the weight at that point. Moreover, the measurement operators must satisfy Eq. (1).

Consider a set of particles, where each particle \( n \) has a mass \( q_n \) and is located at a point \( R_n \) in the allowed region pertaining to a single outcome. Without a loss of generality, the total mass can be assumed to be \( \sum_n q_n = 1 \).

By definition, there exists a measurement operator \( \hat{M}_n \) with singular values \( \{\lambda_{ni}\} \) that corresponds to the point \( R_n \). In general, its weight \( p(n) = \sigma_n^2 / d \) is not equal to the mass \( q_n \). However, the weight can be adjusted by rescaling and duplicating \( \hat{M}_n \). That is, for a particle \( n \), \( d \) measurement operators are introduced such that

\[
\hat{M}_{ns} \equiv \sqrt{\frac{q_n}{\sigma_n^2}} \sum_i \lambda_{ni} |c_s(i)\rangle \langle c_s(i)|
\]

with \( s = 0, 1, \ldots, d-1 \), where \( c_s(i) \equiv (i-1+s \mod d)+1 \) performs the cyclic permutation of \( \{|i\rangle\} \). These measurement operators correspond to the same point \( R_n \) from the interchange invariance in Eq. (30), giving the same weight \( q_n / d \). Note that the weight is not invariant under rescaling of the singular values in Eq. (31). The total weight of the \( d \) measurement operators is then equal to the mass \( q_n \) as desired. Moreover, such measurement operators for all the particles satisfy Eq. (1) such that \( \sum_{n,s} \hat{M}_{ns}^\dagger \hat{M}_{ns} = \sum_n q_n I = \hat{I} \) when regarding a pair of indices \( (n,s) \) as an outcome \( m \). Therefore, \( \{\hat{M}_{ns}\} \) is a measurement equivalent to the set of particles.

In this construction, one particle corresponds to \( d \) outcomes, even though the number of outcomes can be reduced when some singular values are degenerate. As a result, it suffices to consider measurements having at most \( 2d \) outcomes to study the allowed regions for the average values because for any point in the region there exists a set of two particles whose center of mass is located at that point.

References

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[2] C. H. Bennett and G. Brassard, in Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India (IEEE, New York, 1984), pp. 175–179.

[3] A. K. Ekert, Phys. Rev. Lett. 67, 661 (1991).

[4] C. H. Bennett, Phys. Rev. Lett. 68, 3121 (1992).

[5] C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. 68, 557 (1992).

[6] C. A. Fuchs and A. Peres, Phys. Rev. A 53, 2038 (1996).

[7] K. Banaszek, Phys. Rev. Lett. 86, 1366 (2001).

[8] C. A. Fuchs and K. Jacobs, Phys. Rev. A 63, 062305 (2001).

[9] K. Banaszek and I. Devetak, Phys. Rev. A 64, 052307 (2001).

[10] H. Barnum, arXiv:quant-ph/0205155.

[11] G. M. D’Ariano, Fortschr. Phys. 51, 318 (2003).

[12] M. Ozawa, Ann. Phys. (NY) 311, 350 (2004).

[13] M. G. Genoni and M. G. A. Paris, Phys. Rev. A 71, 052307 (2005).

[14] L. Mišta, Jr., J. Fiurášek, and R. Filip, Phys. Rev. A 72, 012311 (2005).

[15] L. Maccone, Phys. Rev. A 73, 042307 (2006).

[16] M. F. Sacchi, Phys. Rev. Lett. 96, 220502 (2006).

[17] F. Buscemi and M. F. Sacchi, Phys. Rev. A 74, 052320 (2006).

[18] K. Banaszek, Open Syst. Inf. Dyn. 13, 1 (2006).

[19] F. Buscemi, M. Hayashi, and M. Horodecki, Phys. Rev. Lett. 100, 210504 (2008).

[20] X.-J. Ren and H. Fan, J. Phys. A: Math. Theor. 47, 305302 (2014).

[21] L. Fan, W. Ge, H. Nha, and M. S. Zubairy, Phys. Rev. A 92, 022114 (2015).
[22] T. Shitara, Y. Kuramochi, and M. Ueda, Phys. Rev. A 93, 032134 (2016).

[23] H. Terashima, Phys. Rev. A 83, 032111 (2011).

[24] H. Terashima, Phys. Rev. A 83, 032114 (2011).

[25] Y. W. Cheong and S.-W. Lee, Phys. Rev. Lett. 109, 150402 (2012).

[26] H. Terashima, Phys. Rev. A 93, 022104 (2016).

[27] M. Ueda and M. Kitagawa, Phys. Rev. Lett. 68, 3424 (1992).

[28] M. Ueda, N. Imoto, and H. Nagaoka, Phys. Rev. A 53, 3808 (1996).

[29] M. Ueda, in Frontiers in Quantum Physics: Proceedings of the International Conference on Frontiers in Quantum Physics, Kuala Lumpur, Malaysia, 1997, edited by S. C. Lim, R. Abd-Shukor, and K. H. Kwek (Springer, Singapore, 1998), pp. 136–144.

[30] F. Sciarrino, M. Ricci, F. De Martini, R. Filip, and L. Mišta, Jr., Phys. Rev. Lett. 96, 020408 (2006).

[31] S.-Y. Baek, Y. W. Cheong, and Y.-H. Kim, Phys. Rev. A 77, 060308(R) (2008).

[32] G. Chen, Y. Zou, X.-Y. Xu, J.-S. Tang, Y.-L. Li, J.-S. Xu, Y.-J. Han, C.-F. Li, G.-C. Guo, H.-Q. Ni, Y. Yu, M.-F. Li, G.-W. Zha, Z.-C. Niu, and Y. Kedem, Phys. Rev. X 4, 021043 (2014).

[33] H.-T. Lim, Y.-S. Ra, K.-H. Hong, S.-W. Lee, and Y.-H. Kim, Phys. Rev. Lett. 113, 020504 (2014).

[34] M. Koashi and M. Ueda, Phys. Rev. Lett. 82, 2598 (1999).

[35] A. Royer, Phys. Rev. Lett. 73, 913 (1994); 74, 1040(E) (1995).

[36] M. A. Nielsen and C. M. Caves, Phys. Rev. A 55, 2547 (1997).

[37] H. Terashima and M. Ueda, Phys. Rev. A 81, 012110 (2010).

[38] H. Terashima, Phys. Rev. A 85, 022124 (2012).

[39] M. Avriel, Nonlinear Programming: Analysis and Methods (Dover Publications, New York, 2003).