Supplementary Information for

Simultaneous Measurement of Multiple Incompatible Observables and Tradeoff in Multiparameter Quantum Estimation

Hongzhen Chen,[1,2,3,*] Lingna Wang,[3,†] and Haidong Yuan[3,‡]

[1] *Institute of Quantum Precision Measurement, State Key Laboratory of Radio Frequency Heterogeneous Integration, College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen, China*
[2] *Quantum Science Center of Guangdong-Hong Kong-Macao Greater Bay Area (Guangdong), Shenzhen, China*
[3] *Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China*

S1. ANALYTICAL ERROR-TRADEOFF RELATION FOR MULTIPLE OBSERVABLES

In this section we derive the tradeoff relation

$$\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}}) \geq \left(\sqrt{\|S_{\mathrm{Re}}^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}}S_{\mathrm{Re}}^{-\frac{1}{2}}\|_F+1}-1\right)^2. \tag{S1}$$

Recall given the observables $\{X_j|1\leq j\leq n\}$, we can construct $\{F_j=\sum_m f_j(m)V_m|1\leq j\leq n\}$ to approximate the observables from a set of projective measurement $\{V_m\}$ on the extended space with an acilla. The mean squared error of the approximation is given by

$$\epsilon_j^2=\mathrm{Tr}\left[(F_j-X_j\otimes I)^2(\rho\otimes\sigma)\right], \tag{S2}$$

where $\sigma=|\xi_0\rangle\langle\xi_0|$ is a state of the ancilla. We then let

$$\begin{aligned}\mathcal{A}_u&=\begin{pmatrix}\langle u|\sqrt{\rho\otimes\sigma}E_1\\ \vdots\\ \langle u|\sqrt{\rho\otimes\sigma}E_n\\ \langle u|\sqrt{\rho\otimes\sigma}(X_1\otimes I)\\ \vdots\\ \langle u|\sqrt{\rho\otimes\sigma}(X_n\otimes I)\end{pmatrix}\begin{pmatrix}\langle u|\sqrt{\rho\otimes\sigma}E_1\\ \vdots\\ \langle u|\sqrt{\rho\otimes\sigma}E_n\\ \langle u|\sqrt{\rho\otimes\sigma}(X_1\otimes I)\\ \vdots\\ \langle u|\sqrt{\rho\otimes\sigma}(X_n\otimes I)\end{pmatrix}^\dagger\\ &=\begin{pmatrix}Q_u & R_u\\ R_u^\dagger & S_u\end{pmatrix}\geq 0,\end{aligned} \tag{S3}$$

* hzchen@szu.edu.cn
† lnwang@mae.cuhk.edu.hk
‡ hdyuan@mae.cuhk.edu.hk

here $E_j = F_j - X_j \otimes I$, $|u\rangle$ is any vector, Q_u, R_u, S_u are $n \times n$ matrices with the entries given by (note E_j and X_j are all Hermitian operators)

$$\begin{aligned}(Q_u)_{jk} &= \langle u|\sqrt{\rho\otimes\sigma}E_jE_k\sqrt{\rho\otimes\sigma}|u\rangle,\\ (R_u)_{jk} &= \langle u|\sqrt{\rho\otimes\sigma}E_j(X_k\otimes I)\sqrt{\rho\otimes\sigma}|u\rangle,\\ (S_u)_{jk} &= \langle u|\sqrt{\rho\otimes\sigma}(X_j\otimes I)(X_k\otimes I)\sqrt{\rho\otimes\sigma}|u\rangle.\end{aligned} \tag{S4}$$

We can write these matrices in terms of the real and imaginary parts as $Q_u = Q_{u,\mathrm{Re}} + iQ_{u,\mathrm{Im}}$, $R_u = R_{u,\mathrm{Re}} + iR_{u,\mathrm{Im}}$, $S_u = S_{u,\mathrm{Re}} + iS_{u,\mathrm{Im}}$, where

$$\begin{aligned}(Q_{u,\mathrm{Re}})_{jk} &= \frac{1}{2}\langle u|\sqrt{\rho\otimes\sigma}\{E_j, E_k\}\sqrt{\rho\otimes\sigma}|u\rangle,\\ (R_{u,\mathrm{Re}})_{jk} &= \frac{1}{2}\langle u|\sqrt{\rho\otimes\sigma}\{E_j, X_k\otimes I\}\sqrt{\rho\otimes\sigma}|u\rangle,\\ (S_{u,\mathrm{Re}})_{jk} &= \frac{1}{2}\langle u|\sqrt{\rho\otimes\sigma}\{X_j\otimes I, X_k\otimes I\}\sqrt{\rho\otimes\sigma}|u\rangle,\end{aligned} \tag{S5}$$

and

$$\begin{aligned}(Q_{u,\mathrm{Im}})_{jk} &= \frac{1}{2i}\langle u|\sqrt{\rho\otimes\sigma}[E_j, E_k]\sqrt{\rho\otimes\sigma}|u\rangle,\\ (R_{u,\mathrm{Im}})_{jk} &= \frac{1}{2i}\langle u|\sqrt{\rho\otimes\sigma}[E_j, X_k\otimes I]\sqrt{\rho\otimes\sigma}|u\rangle,\\ (S_{u,\mathrm{Im}})_{jk} &= \frac{1}{2i}\langle u|\sqrt{\rho\otimes\sigma}[X_j\otimes I, X_k\otimes I]\sqrt{\rho\otimes\sigma}|u\rangle.\end{aligned} \tag{S6}$$

For any set of $\{|u_q\rangle\}$ with $\sum_q |u_q\rangle\langle u_q| = I$, we can obtain a corresponding set of $\{\mathcal{A}_{u_q}\}$. We then let $\tilde{\mathcal{A}} = \sum_q \tilde{\mathcal{A}}_{u_q}$ with each $\tilde{\mathcal{A}}_{u_q}$ equals to either $\mathcal{A}_{u_q}$ or $\mathcal{A}_{u_q}^T$. Since $\mathcal{A}_{u_q} \geq 0$, and $\mathcal{A}_{u_q}^T \geq 0$, we have

$$\tilde{\mathcal{A}} = \sum_q \tilde{\mathcal{A}}_{u_q} = \begin{pmatrix}\tilde{Q} & \tilde{R}\\ \tilde{R}^\dagger & \tilde{S}\end{pmatrix} \geq 0, \tag{S7}$$

where $\tilde{Q} = \sum_q \tilde{Q}_{u_q}$, $\tilde{R} = \sum_q \tilde{R}_{u_q}$, $\tilde{S} = \sum_q \tilde{S}_{u_q}$ with each $(\tilde{Q}_{u_q}, \tilde{R}_{u_q}, \tilde{S}_{u_q})$ equals to either $(Q_{u_q}, R_{u_q}, S_{u_q})$ or $(\bar{Q}_{u_q}, \bar{R}_{u_q}, \bar{S}_{u_q})$, here $\bar{M} = M_{\mathrm{Re}} - iM_{\mathrm{Im}}$. We assume $\tilde{S}$ is nonsingular, which means $\{X_j\}$ are linearly independent.

Since $\tilde{\mathcal{A}} \geq 0$, with the Schur complement, we have

$$\tilde{Q} - \tilde{R}\tilde{S}^{-1}\tilde{R}^\dagger \geq 0. \tag{S8}$$

This can be rewritten as $\tilde{Q} - (i\tilde{R})\tilde{S}^{-1}(i\tilde{R})^\dagger \geq 0$, which is the Schur complement of

$$\tilde{\mathcal{B}} = \begin{pmatrix}\tilde{Q} & i\tilde{R}\\ (i\tilde{R})^\dagger & \tilde{S}\end{pmatrix}. \tag{S9}$$

Since $\tilde{Q} - (i\tilde{R})\tilde{S}^{-1}(i\tilde{R})^\dagger \geq 0$ and $\tilde{S} > 0$, we then have $\tilde{\mathcal{B}} \geq 0$. This implies $\tilde{\mathcal{B}}_{\mathrm{Re}} = \frac{1}{2}(\tilde{\mathcal{B}} + \tilde{\mathcal{B}}^T) \geq 0$, which can be written as

$$\tilde{\mathcal{B}}_{\mathrm{Re}} = \begin{pmatrix}\tilde{Q}_{\mathrm{Re}} & -\tilde{R}_{\mathrm{Im}}\\ -\tilde{R}_{\mathrm{Im}}^T & \tilde{S}_{\mathrm{Re}}\end{pmatrix} \geq 0. \tag{S10}$$

With the Schur complement of $\tilde{\mathcal{B}}_{\mathrm{Re}}$ we have

$$\tilde{Q}_{\mathrm{Re}} - \tilde{R}_{\mathrm{Im}}\tilde{S}_{\mathrm{Re}}^{-1}\tilde{R}_{\mathrm{Im}}^T \geq 0. \tag{S11}$$

This implies

$$\tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}}\tilde{Q}_{\mathrm{Re}}\tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}} - \tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}}\tilde{R}_{\mathrm{Im}}\tilde{S}_{\mathrm{Re}}^{-1}\tilde{R}_{\mathrm{Im}}^T\tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}} \geq 0, \tag{S12}$$

thus

$$\begin{aligned}\mathrm{Tr}(\tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}}\tilde{Q}_{\mathrm{Re}}\tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}}) &\geq \mathrm{Tr}(\tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}}\tilde{R}_{\mathrm{Im}}\tilde{S}_{\mathrm{Re}}^{-1}\tilde{R}_{\mathrm{Im}}^T\tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}})\\ &= \|\tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}}\tilde{R}_{\mathrm{Im}}\tilde{S}_{\mathrm{Re}}^{-\frac{1}{2}}\|_F^2,\end{aligned} \tag{S13}$$

where $\|\cdot\|_F$ is the Frobenius norm. Here the right side of the inequality depends on $\tilde{R}_{\rm Im}$, which depends on the choice of the measurement. To get rid of $\tilde{R}_{\rm Im}$, we use the fact that $[F_j, F_k] = 0$ since they are constructed from the same projective measurement. $[F_j, F_k] = 0$ can be rewritten as $[E_j + X_j \otimes I, E_k + X_k \otimes I] = 0$, which can be expanded as

$$[E_j, E_k] + [E_j, X_k \otimes I] + [X_j \otimes I, E_k] + [X_j \otimes I, X_k \otimes I] = 0. \tag{S14}$$

From which we have $Q_{u_q,\rm Im} + R_{u_q,\rm Im} - R^T_{u_q,\rm Im} + S_{u_q,\rm Im} = 0$. Since $(\tilde{Q}_{u_q}, \tilde{R}_{u_q}, \tilde{S}_{u_q})$ equal to either $(Q_{u_q}, R_{u_q}, S_{u_q})$ or $(\bar{Q}_{u_q}, \bar{R}_{u_q}, \bar{S}_{u_q})$, $(\tilde{Q}_{u_q,\rm Im}, \tilde{R}_{u_q,\rm Im}, \tilde{S}_{u_q,\rm Im})$ then equal to either $(Q_{u_q,\rm Im}, R_{u_q,\rm Im}, S_{u_q,\rm Im})$ or $(-Q_{u_q,\rm Im}, -R_{u_q,\rm Im}, -S_{u_q,\rm Im})$. In either case $\tilde{Q}_{u_q,\rm Im} + \tilde{R}_{u_q,\rm Im} - \tilde{R}^T_{u_q,\rm Im} + \tilde{S}_{u_q,\rm Im} = 0$ holds, thus $\sum_q(\tilde{Q}_{u_q,\rm Im} + \tilde{R}_{u_q,\rm Im} - \tilde{R}^T_{u_q,\rm Im} + \tilde{S}_{u_q,\rm Im}) = 0$. This can be written as

$$\tilde{Q}_{\rm Im} + \tilde{R}_{\rm Im} - \tilde{R}^T_{\rm Im} + \tilde{S}_{\rm Im} = 0. \tag{S15}$$

By multiplying $\tilde{S}_{\rm Re}^{-\frac{1}{2}}$ from both left and right, we have $\hat{Q}_{\rm Im} + \hat{R}_{\rm Im} - \hat{R}^T_{\rm Im} + \hat{S}_{\rm Im} = 0$, which can be equivalently written as

$$\hat{S}_{\rm Im} = -\hat{Q}_{\rm Im} - \hat{R}_{\rm Im} + \hat{R}^T_{\rm Im}, \tag{S16}$$

where $\hat{\square} = \tilde{S}_{\rm Re}^{-\frac{1}{2}} \tilde{\square} \tilde{S}_{\rm Re}^{-\frac{1}{2}}$. Applying the triangle inequality we get

$$\|\hat{S}_{\rm Im}\|_F \leq \|\hat{Q}_{\rm Im}\|_F + \|\hat{R}_{\rm Im}\|_F + \|\hat{R}^T_{\rm Im}\|_F. \tag{S17}$$

Note that Eq.(S13) can be rewritten as $\|\hat{R}^T_{\rm Im}\|_F^2 = \|\hat{R}_{\rm Im}\|_F^2 \leq \mathrm{Tr}(\hat{Q}_{\rm Re})$, and from $\hat{Q} \geq 0$ we have

$$\mathrm{Tr}(\hat{Q}_{\rm Re}) \geq \|\hat{Q}_{\rm Im}\|_1 \geq \|\hat{Q}_{\rm Im}\|_F. \tag{S18}$$

Combining these inequalities we then have

$$\begin{aligned}\|\hat{S}_{\rm Im}\|_F &\leq \|\hat{Q}_{\rm Im}\|_F + \|\hat{R}_{\rm Im}\|_F + \|\hat{R}^T_{\rm Im}\|_F \\ &\leq \mathrm{Tr}(\hat{Q}_{\rm Re}) + 2\sqrt{\mathrm{Tr}(\hat{Q}_{\rm Re})}.\end{aligned} \tag{S19}$$

This leads to the bound on $\mathrm{Tr}(\hat{Q}_{\rm Re})$ as

$$\mathrm{Tr}(\hat{Q}_{\rm Re}) \geq \left(\sqrt{\|\hat{S}_{\rm Im}\|_F + 1} - 1\right)^2, \tag{S20}$$

which can be equivalently written as

$$\mathrm{Tr}(S_{\rm Re}^{-1} Q_{\rm Re}) \geq \left(\sqrt{\|S_{\rm Re}^{-\frac{1}{2}} \tilde{S}_{\rm Im} S_{\rm Re}^{-\frac{1}{2}}\|_F + 1} - 1\right)^2. \tag{S21}$$

We note that this tradeoff relation is invariant under linear transformations of the observables, i.e., if we let $Y_j = \sum_k b_{jk} X_k$ with b_{jk} as the jk-th entry of a non-singular real matrix B, then $Q_{\rm Re}(Y) = B^T Q_{\rm Re}(X) B$, $\tilde{S}(Y) = B^T \tilde{S}(x) B$, we then have $\mathrm{Tr}[S_{\rm Re}^{-1}(Y) Q_{\rm Re}(Y)] = \mathrm{Tr}[S_{\rm Re}^{-1}(X) Q_{\rm Re}(X)]$ and $\|S_{\rm Re}^{-\frac{1}{2}}(Y) \tilde{S}_{\rm Im}(Y) S_{\rm Re}^{-\frac{1}{2}}(Y)\|_F = \|S_{\rm Re}^{-\frac{1}{2}}(X) \tilde{S}_{\rm Im}(X) S_{\rm Re}^{-\frac{1}{2}}(X)\|_F$. The tradeoff relation thus remain as the same under linear transformation and it is often convenient to choose a linear transformation under which $S_{\rm Re} = I$.

S2. ERROR-TRADEOFF RELATION WITH SEMIDEFINITE PROGRAMMING

Recall that given a set of observables $\{X_j\}_{j=1}^n$ in Hilbert space $\mathcal{H}_S$, we can construct $\{F_j = \sum_m f_j(m) V_m\}_{j=1}^n$ to approximate the observables from a set of projective measurement $\{V_m\}$ on the extended space $\mathcal{H}_S \otimes \mathcal{H}_A$. We denote the weighted mean squared error of the approximations as

$$\mathcal{E} = \sum_{j=1}^n w_j \epsilon_j^2 = \sum_{j=1}^n w_j \mathrm{Tr}[(\rho \otimes \sigma)(F_j - X_j \otimes I)^2], \tag{S22}$$

where $\rho \in \mathcal{H}_S$ is the state of the system and $\sigma = |\xi_0\rangle\langle\xi_0| \in \mathcal{H}_A$ is the state of the ancilla. To formularize the minimization of $\mathcal{E}$ over all choices of $\{F_j\}_{j=1}^n$, we first define an $n \times n$ Hermitian matrix Q, whose jk-th element is given by

$$
\begin{aligned}
Q_{jk} =& \mathrm{Tr}\left[(\rho\otimes\sigma)(F_j - X_j\otimes I)(F_k - X_k\otimes I)\right] \\
=& \mathrm{Tr}\left[(\rho\otimes\sigma)F_jF_k\right] - \mathrm{Tr}\left[(\rho\otimes\sigma)F_j(X_k\otimes I)\right] - \mathrm{Tr}\left[(\rho\otimes\sigma)(X_j\otimes I)F_k\right] + \mathrm{Tr}\left[(\rho\otimes\sigma)(X_j\otimes I)(X_k\otimes I)\right] \\
=& \mathrm{Tr}\left[\rho\sum_m f_j(m)M_mf_k(m)\right] - \mathrm{Tr}\left(\rho R_jX_k\right) - \mathrm{Tr}\left(\rho X_jR_k\right) + \mathrm{Tr}\left(\rho X_jX_k\right),
\end{aligned}
\tag{S23}
$$

where $R_j = (I\otimes\langle\xi_0|)F_j(I\otimes|\xi_0\rangle) = \sum_m f_j(m)M_m$ for $1 \le j \le n$ are Hermitian matrices in $\mathcal{H}_S$, $\{M_m = (I\otimes\langle\xi_0|)V_m(I\otimes|\xi_0\rangle)\}$ forms a POVM. We denote $\mathbb{S}_{jk} = \sum_m f_j(m)M_mf_k(m)$, $\mathbb{R} = \begin{pmatrix}R_1 & R_2 & \cdots & R_n\end{pmatrix}^\dagger$ and $\mathbb{X} = \begin{pmatrix}X_1 & X_2 & \cdots & X_n\end{pmatrix}^\dagger$, the Hermitian matrix Q can then be written as $Q = \mathrm{Tr}_S\left[(I_n\otimes\rho)(\mathbb{S} - \mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^\dagger + \mathbb{X}\mathbb{X}^\dagger)\right]$, and

$$
\mathcal{E} = \mathrm{Tr}(WQ) = \mathrm{Tr}\left[(W\otimes\rho)(\mathbb{S} - \mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^\dagger + \mathbb{X}\mathbb{X}^\dagger)\right],
\tag{S24}
$$

where $\mathbb{S}$ is an $n\times n$ block matrix with the jk-th block as $\mathbb{S}_{jk}$, which is itself a Hermitian operator, $W \ge 0$ is a weight matrix, which is typically taken as $W = \mathrm{diag}\{w_1, \cdots, w_n\}$, but it can also be taken as a general positive semidefinite matrix.

We now show that for the defined $\mathbb{S}$ and $\mathbb{R}$, we have $\mathbb{S} \ge \mathbb{R}\mathbb{R}^\dagger$. Since $\mathbb{S}$ and $\mathbb{R}\mathbb{R}^\dagger$ are $nd_s \times nd_s$ matrices, where d_S is the dimension of $\mathcal{H}_S$, $\mathbb{S} \ge \mathbb{R}\mathbb{R}^\dagger$ is equivalent to $b^\dagger(\mathbb{S} - \mathbb{R}\mathbb{R}^\dagger)b \ge 0$ for any $nd_S \times 1$ vector b.

For any $nd_S \times 1$ vector, b, we can write $b^\dagger$ as $b^\dagger = \begin{pmatrix}\langle b_1| & \langle b_2| & \cdots & \langle b_n|\end{pmatrix}$ with $\{\langle b_j|\}_{j=1}^n$ being d_S-dimensional row vectors. We then have

$$
\begin{aligned}
& b^\dagger\mathbb{S}b - b^\dagger\mathbb{R}\mathbb{R}^\dagger b = \sum_{j,k}\langle b_j|(\mathbb{S}_{jk} - R_jR_k)|b_k\rangle \\
=& \sum_{j,k}\langle b_j|\left[\sum_m f_j(m)M_mf_k(m) - \left(\sum_m f_j(m)M_m\right)\left(\sum_l f_k(l)M_l\right)\right]|b_k\rangle \\
=& \sum_m\left(\sum_j\langle b_j|f_j(m)\right)M_m\left(\sum_k f_k(m)|b_k\rangle\right) - \left[\sum_m\left(\sum_j\langle b_j|f_j(m)\right)M_m\right]\left[\sum_l M_l\left(\sum_k f_k(l)|b_k\rangle\right)\right] \\
=& \sum_m\left[\sum_j\langle b_j|f_j(m) - \sum_l\left(\sum_j\langle b_j|f_j(l)\right)M_l\right]M_m\left[\sum_k f_k(m)|b_k\rangle - \sum_l M_l\left(\sum_k f_k(l)|b_k\rangle\right)\right] \\
=& \sum_m V_m^\dagger M_mV_m \ge 0,
\end{aligned}
\tag{S25}
$$

where $V_m = \sum_k f_k(m)|b_k\rangle - \sum_l M_l\left(\sum_k f_k(l)|b_k\rangle\right) = \sum_j f_j(m)|b_j\rangle - \sum_l M_l\left(\sum_j f_j(l)|b_j\rangle\right)$. We thus have $\mathbb{S} \ge \mathbb{R}\mathbb{R}^\dagger$.

Denote $\{R_j^\star\}_{j=1}^n$ and $\mathbb{S}^\star$ as the optimal operators that gives the minimal error, $\mathcal{E}^\star$, we then have

$$
\begin{aligned}
\mathcal{E} \ge \mathcal{E}^\star &= \mathrm{Tr}\left[(W\otimes\rho)(\mathbb{S}^\star - \mathbb{R}^\star\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^{\star\dagger} + \mathbb{X}\mathbb{X}^\dagger)\right] \\
&\ge \min_{\mathbb{S}}\left\{\mathrm{Tr}\left[(W\otimes\rho)(\mathbb{S} - \mathbb{R}^\star\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^{\star\dagger} + \mathbb{X}\mathbb{X}^\dagger)\right] | \mathbb{S}_{jk} = \mathbb{S}_{jk}^\dagger = \mathbb{S}_{kj}, \mathbb{S} \ge \mathbb{R}^\star\mathbb{R}^{\star\dagger}\right\} \\
&\ge \min_{\mathbb{S},\{R_j\}_{j=1}^n}\left\{\mathrm{Tr}\left[(W\otimes\rho)(\mathbb{S} - \mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^\dagger + \mathbb{X}\mathbb{X}^\dagger)\right] | \mathbb{S}_{jk} = \mathbb{S}_{jk}^\dagger = \mathbb{S}_{kj}, \mathbb{S} \ge \mathbb{R}\mathbb{R}^\dagger, R_j = R_j^\dagger\right\}.
\end{aligned}
\tag{S26}
$$

And the minimization can be formulated as a semidefinite programming as

$$
\begin{aligned}
\mathcal{E}_0 = & \min_{\mathbb{S},\{R_j\}_{j=1}^n} \mathrm{Tr}\left[(W\otimes\rho)(\mathbb{S} - \mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^\dagger + \mathbb{X}\mathbb{X}^\dagger)\right] \\
& \text{subject to} \quad \mathbb{S}_{jk} = \mathbb{S}_{kj} = \mathbb{S}_{jk}^\dagger, \ \forall j,k \\
& \qquad\qquad\quad R_j = R_j^\dagger, \ \forall j \\
& \qquad\qquad\quad \begin{pmatrix} I & \mathbb{R}^\dagger \\ \mathbb{R} & \mathbb{S}\end{pmatrix} \ge 0.
\end{aligned}
\tag{S27}
$$

S3. COMPARISON OF THE SDP AND ANALYTICAL BOUND

In this section we show that the bound $\mathcal{E}_0$ is tighter than the analytical bounds with any choice of $\{|u_q\rangle\}$.

Recall that for any vector $|u\rangle$ in $\mathcal{H}_S \otimes \mathcal{H}_A$, we have

$$\begin{aligned}\mathcal{A}_u &= \begin{pmatrix} \langle u|\sqrt{\rho\otimes\sigma}(F_1 - X_1\otimes I) \\ \vdots \\ \langle u|\sqrt{\rho\otimes\sigma}(F_n - X_n\otimes I) \\ \langle u|\sqrt{\rho\otimes\sigma}(X_1\otimes I) \\ \vdots \\ \langle u|\sqrt{\rho\otimes\sigma}(X_n\otimes I)\end{pmatrix}\begin{pmatrix} \langle u|\sqrt{\rho\otimes\sigma}(F_1 - X_1\otimes I) \\ \vdots \\ \langle u|\sqrt{\rho\otimes\sigma}(F_n - X_n\otimes I) \\ \langle u|\sqrt{\rho\otimes\sigma}(X_1\otimes I) \\ \vdots \\ \langle u|\sqrt{\rho\otimes\sigma}(X_n\otimes I)\end{pmatrix}^\dagger \\ &= \begin{pmatrix} Q_u & D_u \\ D_u^\dagger & J_u\end{pmatrix} \geq 0,\end{aligned} \tag{S28}$$

where Q_u, D_u, J_u are $n\times n$ matrices with the jk-th elements given by

$$\begin{aligned}(Q_u)_{jk} &= \langle u|\sqrt{\rho\otimes\sigma}(F_j - X_j\otimes I)(F_k - X_k\otimes I)\sqrt{\rho\otimes\sigma}|u\rangle, \\ (D_u)_{jk} &= \langle u|\sqrt{\rho\otimes\sigma}(F_j - X_j\otimes I)(X_k\otimes I)\sqrt{\rho\otimes\sigma}|u\rangle, \\ (J_u)_{jk} &= \langle u|\sqrt{\rho\otimes\sigma}(X_j\otimes I)(X_k\otimes I)\sqrt{\rho\otimes\sigma}|u\rangle.\end{aligned} \tag{S29}$$

Note that $[F_j, F_k] = 0$, which gives

$$(Q_u + D_u + D_u^\dagger + J_u)_{jk} = (Q_u + D_u + D_u^\dagger + J_u)_{kj}, \tag{S30}$$

i.e., the matrix $Q_u + D_u + D_u^\dagger + J_u$ is real symmetric. For any set of vectors $\{|u_q\rangle\}$ that satistifies $\sum_q |u_q\rangle\langle u_q| = I$, we have $Q = \sum_q Q_{u_q}$. Together with Eq.(S28) and Eq.(S30), we can write the bound as a minimization,

$$\begin{aligned}\mathcal{E}_u = \min_{\{F_j\}} \;& \mathrm{Tr}[WQ] = \sum_q \mathrm{Tr}[WQ_{u_q}] \\ \text{subject to} \;& \begin{pmatrix} Q_{u_q} & D_{u_q} \\ D_{u_q}^\dagger & J_{u_q}\end{pmatrix} \geq 0, \forall q \\ & Q_{u_q} + D_{u_q} + D_{u_q}^\dagger + J_{u_q} \text{ real symmetric}, \forall q.\end{aligned} \tag{S31}$$

$\mathcal{E}_u$ is tighter than any analytical bounds based on it, unless any other potential constraints could be introduced. In particular it is tighter than the analytical bound obtained in Appendix S1. We will then show $\mathcal{E}_0$ is tighter than $\mathcal{E}_u$ for any $\{|u_q\rangle\}$.

We first formulate $\mathcal{E}_u$ as a semidefinite programming for a fixed $\{|u_q\rangle\}$, then show that $\mathcal{E}_0 \geq \mathcal{E}_u$ for any $\{|u_q\rangle\}$. First note that since $\sigma = |\xi_0\rangle\langle\xi_0|$, we have

$$\begin{aligned}&\sqrt{\rho\otimes\sigma}|u_q\rangle\langle u_q|\sqrt{\rho\otimes\sigma} \\ &= (\sqrt{\rho}\otimes|\xi_0\rangle)(I\otimes\langle\xi_0|)|u_q\rangle\langle u_q|(I\otimes|\xi_0\rangle)(\sqrt{\rho}\otimes\langle\xi_0|) \\ &= (\sqrt{\rho}\otimes|\xi_0\rangle)\tilde{\rho}(\sqrt{\rho}\otimes\langle\xi_0|) \\ &= \sqrt{\rho}\tilde{\rho}\sqrt{\rho}\otimes|\xi_0\rangle\langle\xi_0| \\ &= \rho_{u_q}\otimes\sigma,\end{aligned} \tag{S32}$$

here $\tilde{\rho} = (I\otimes\langle\xi_0|)|u_q\rangle\langle u_q|(I\otimes|\xi_0\rangle)$ and $\rho_{u_q} = \sqrt{\rho}\tilde{\rho}\sqrt{\rho}$. We note that $\sum_q \rho_{u_q} = \rho$, and each ρ_{u_q} is unnormalized with rank equal to 1. With $\mathbb{S}_{jk} = \sum_m f_j(m)M_m f_k(m)$, $\mathbb{R} = \begin{pmatrix} R_1 & R_2 & \cdots & R_n\end{pmatrix}^\dagger$ and $\mathbb{X} = \begin{pmatrix} X_1 & X_2 & \cdots & X_n\end{pmatrix}^\dagger$, we have $Q_{u_q} = \mathrm{Tr}_S\left[(I_n\otimes\rho_{u_q})(\mathbb{S} - \mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^\dagger + \mathbb{X}\mathbb{X}^\dagger)\right]$, $D_{u_q} = \mathrm{Tr}_S\left[(I_n\otimes\rho_{u_q})(\mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{X}^\dagger)\right]$ and $J_{u_q} = \mathrm{Tr}_S\left[(I_n\otimes\rho_{u_q})\mathbb{X}\mathbb{X}^\dagger\right]$.

With Schur complement, $\begin{pmatrix} Q_{u_q} & D_{u_q} \\ D_{u_q}^\dagger & J_{u_q} \end{pmatrix} \geq 0$ is equivalent to $Q_{u_q} \geq D_{u_q} J_{u_q}^{-1} D_{u_q}^\dagger$, which further gives

$$\begin{aligned}
& Q_{u_q} + (D_{u_q} + J_{u_q}) + (D_{u_q} + J_{u_q})^\dagger - J_{u_q} \geq (D_{u_q} + J_{u_q}) J_{u_q}^{-1} (D_{u_q} + J_{u_q})^\dagger \\
\Leftrightarrow\quad & \mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{S}\right] \geq \mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{R}\mathbb{X}^\dagger\right] \mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{X}\mathbb{X}^\dagger\right]^{-1} \mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{R}\mathbb{X}^\dagger\right]^\dagger \\
\Leftrightarrow\quad & \begin{pmatrix} \mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{S}\right] & \mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{R}\mathbb{X}^\dagger\right] \\ \mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{R}\mathbb{X}^\dagger\right]^\dagger & \mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{X}\mathbb{X}^\dagger\right] \end{pmatrix} \geq 0 \\
\Leftrightarrow\quad & \mathrm{Tr}_S\left[(I_2 \otimes I_n \otimes \rho_{u_q}) \begin{pmatrix} \mathbb{S} & \mathbb{R}\mathbb{X}^\dagger \\ \mathbb{X}\mathbb{R}^\dagger & \mathbb{X}\mathbb{X}^\dagger \end{pmatrix}\right] \geq 0.
\end{aligned} \tag{S33}$$

Also note that the constraint of $Q_{u_q} + D_{u_q} + D_{u_q}^\dagger + J_{u_q}$ being real symmetric is equivalent to $\mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{S}\right]$ being real symmetric. $\mathcal{E}_u$ can then be formulated as a semidefinite programming for a fixed $\{|u_q\rangle\}$ as

$$\begin{aligned}
\mathcal{E}_u = & \min_{\mathbb{S}, \{R_j\}_{j=1}^n} \mathrm{Tr}\left[(W \otimes \rho)(\mathbb{S} - \mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^\dagger + \mathbb{X}\mathbb{X}^\dagger)\right] \\
& \text{subject to} \quad \mathrm{Tr}_S\left[(I_2 \otimes I_n \otimes \rho_{u_q}) \begin{pmatrix} \mathbb{S} & \mathbb{R}\mathbb{X}^\dagger \\ \mathbb{X}\mathbb{R}^\dagger & \mathbb{X}\mathbb{X}^\dagger \end{pmatrix}\right] \geq 0, \forall q \\
& \qquad\qquad\quad \mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{S}\right] \ \text{real symmetric}, \forall q \\
& \qquad\qquad\quad R_j = R_j^\dagger.
\end{aligned} \tag{S34}$$

Next we show that $\mathcal{E}_0$ is tighter than $\mathcal{E}_u$ for any $\{|u_q\rangle\}$. Since $\begin{pmatrix} \mathbb{R}\mathbb{R}^\dagger & \mathbb{R}\mathbb{X}^\dagger \\ \mathbb{X}\mathbb{R}^\dagger & \mathbb{X}\mathbb{X}^\dagger \end{pmatrix} = \begin{pmatrix} \mathbb{R} \\ \mathbb{X} \end{pmatrix} \begin{pmatrix} \mathbb{R} \\ \mathbb{X} \end{pmatrix}^\dagger \geq 0$, combining with the constraint $\mathbb{S} \geq \mathbb{R}\mathbb{R}^\dagger$ in $\mathcal{E}_0$, we can obtain $\begin{pmatrix} \mathbb{S} & \mathbb{R}\mathbb{X}^\dagger \\ \mathbb{X}\mathbb{R}^\dagger & \mathbb{X}\mathbb{X}^\dagger \end{pmatrix} \geq 0$, which is stronger than the first constraint in $\mathcal{E}_u$. From the constraint $\mathbb{S}_{jk} = \mathbb{S}_{kj} = \mathbb{S}_{jk}^\dagger$, $\forall j, k$ in $\mathcal{E}_0$, we can also obtain that $\mathrm{Tr}_S\left[(I_n \otimes \rho_{u_q})\mathbb{S}\right]$ is real symmetric $\forall q$. The semidefinite programming $\mathcal{E}_0$ thus has the same objective function but stronger constraints than those in $\mathcal{E}_u$, which indicates that $\mathcal{E}_0 \geq \mathcal{E}_u$.

S4. TIGHT ERROR-TRADEOFF RELATION AND OPTIMAL MEASUREMENT FOR PURE STATES

In this section, we show that when ρ is a pure state, $\mathcal{E} \geq \mathcal{E}_0$ is tight. Additionally we provide an explicit construction of the optimal measurement that saturates the bound.

Given any pure state $\rho = |\psi\rangle\langle\psi|$, we have

$$\mathrm{Tr}\left[(W \otimes \rho)(\mathbb{S} - \mathbb{R}\mathbb{X}^\dagger - \mathbb{X}\mathbb{R}^\dagger + \mathbb{X}\mathbb{X}^\dagger)\right] = \mathrm{Tr}\left[W\left(S - R^\dagger X - X^\dagger R + X^\dagger X\right)\right], \tag{S35}$$

where S is an $n \times n$ matrix with its jkth element given as $S_{jk} = \langle\psi|\mathbb{S}_{jk}|\psi\rangle$, $R = \begin{pmatrix} |r_1\rangle & |r_2\rangle & \cdots & |r_n\rangle \end{pmatrix}$ and $X = \begin{pmatrix} |x_1\rangle & |x_2\rangle & \cdots & |x_n\rangle \end{pmatrix}$ are $d_S \times n$ matrices with $|r_j\rangle = R_j|\psi\rangle$, $|x_j\rangle = X_j|\psi\rangle$ for $1 \leq j \leq n$. $\mathbb{S}_{jk} = \mathbb{S}_{kj} = \mathbb{S}_{jk}^\dagger$ indicates that S is real symmetric. From $\mathbb{S} \geq \mathbb{R}\mathbb{R}^\dagger$, we have $S \geq R^\dagger R = (R^\dagger R)_{\mathrm{Re}} + i(R^\dagger R)_{\mathrm{Im}}$. By taking transpose on both sides, we also have $S^T = S \geq (R^\dagger R)_{\mathrm{Re}} - i(R^\dagger R)_{\mathrm{Im}}$, thus $W^{\frac{1}{2}}(S - (R^\dagger R)_{\mathrm{Re}})W^{\frac{1}{2}} \geq \pm i W^{\frac{1}{2}}(R^\dagger R)_{\mathrm{Im}} W^{\frac{1}{2}}$ and we get

$$\mathrm{Tr}(WS) \geq \mathrm{Tr}(W(R^\dagger R)_{\mathrm{Re}}) + \mathrm{Tr}\left|W^{\frac{1}{2}}(R^\dagger R)_{\mathrm{Im}} W^{\frac{1}{2}}\right|, \tag{S36}$$

where the inequality can be saturated if $S = (R^\dagger R)_{\mathrm{Re}} + W^{-\frac{1}{2}}\left|W^{\frac{1}{2}}(R^\dagger R)_{\mathrm{Im}} W^{\frac{1}{2}}\right| W^{-\frac{1}{2}}$. Thus we can further simplify $\mathcal{E}_0$ for pure states as

$$\begin{aligned}
\mathcal{E}_0 &= \min_{S,R} \left\{\mathrm{Tr}\left[W\left(S - R^\dagger X - X^\dagger R + X^\dagger X\right)\right] | S \ \text{real symmetric}, S \geq R^\dagger R\right\} \\
&= \min_R \left\{\mathrm{Tr}\left[W\left((R^\dagger R)_{\mathrm{Re}} - R^\dagger X - X^\dagger R + X^\dagger X\right)\right] + \mathrm{Tr}\left|W^{\frac{1}{2}}(R^\dagger R)_{\mathrm{Im}} W^{\frac{1}{2}}\right|\right\}.
\end{aligned} \tag{S37}$$

The computation of $\mathcal{E}_0$ can be largely simplified by the following lemma.

Lemma 1. $\mathcal{E}_0 = \mathcal{E}_0'$, *where*

$$\mathcal{E}_0' = \min_R \left\{ \mathrm{Tr}\left[W\left((R^\dagger R)_{\mathrm{Re}} - R^\dagger X - X^\dagger R + X^\dagger X \right) \right] | (R^\dagger R)_{\mathrm{Im}} = 0 \right\}. \tag{S38}$$

Proof. Denote $\mathcal{X}$ as the linear span of $\{|\psi\rangle, |x_1\rangle, ..., |x_n\rangle\}$ and $\mathcal{X}^\perp$ as its orthogonal complement. Each $|r_j\rangle$ can then be decomposed as $|r_j\rangle = |a_j\rangle + |b_j\rangle$ with $|a_j\rangle \in \mathcal{X}$ and $|b_j\rangle \in \mathcal{X}^\perp$. Define $A = \begin{pmatrix} |a_1\rangle & |a_2\rangle & \cdots & |a_n\rangle \end{pmatrix}$ and $B = \begin{pmatrix} |b_1\rangle & |b_2\rangle & \cdots & |b_n\rangle \end{pmatrix}$, we then have $R = A + B$ and $R^\dagger R = A^\dagger A + B^\dagger B$. Next we consider the minimization in $\mathcal{E}_0'$ over B with any fixed A. The Lagrangian of the minimization can be written as

$$\begin{aligned} \mathcal{L}(B,\Lambda) &= \mathrm{Tr}\left[W\left((R^\dagger R)_{\mathrm{Re}} - R^\dagger X - X^\dagger R + X^\dagger X \right) \right] + \mathrm{Tr}\left[(R^\dagger R)_{\mathrm{Im}} \Lambda \right] \\ &= \mathrm{Tr}\left[W\left((A^\dagger A)_{\mathrm{Re}} + (B^\dagger B)_{\mathrm{Re}} - A^\dagger X - X^\dagger A + X^\dagger X \right) \right] + \mathrm{Tr}\left[(A^\dagger A)_{\mathrm{Im}} \Lambda \right] + \mathrm{Tr}\left[(B^\dagger B)_{\mathrm{Im}} \Lambda \right]. \end{aligned} \tag{S39}$$

Here since $(R^\dagger R)_{\mathrm{Im}}$ is antisymmetric, we have $\mathrm{Tr}\left[(R^\dagger R)_{\mathrm{Im}} \Lambda \right] = \mathrm{Tr}\left[(R^\dagger R)_{\mathrm{Im}}^T \Lambda^T \right] = -\mathrm{Tr}\left[(R^\dagger R)_{\mathrm{Im}} \Lambda^T \right]$, which further gives $\mathrm{Tr}\left[(R^\dagger R)_{\mathrm{Im}} \Lambda \right] = \mathrm{Tr}\left[(R^\dagger R)_{\mathrm{Im}} (\Lambda - \Lambda^T)/2 \right]$, where $(\Lambda - \Lambda^T)/2$ is the antisymmetric part of Λ. Thus without loss of generality we assume Λ is antisymmetric.

The matrix derivative of $\mathcal{L}(B)$ with respect to B can be calculated as

$$\begin{aligned} \partial_B \mathcal{L}(B,\Lambda) &= \lim_{\varepsilon \to 0} \frac{\mathcal{L}(B + \varepsilon \cdot \delta B, \Lambda) - \mathcal{L}(B,\Lambda)}{\varepsilon} \\ &= \mathrm{Tr}\left[W(B^\dagger \delta B + \delta B^\dagger B)_{\mathrm{Re}} + (B^\dagger \delta B + \delta B^\dagger B)_{\mathrm{Im}} \Lambda \right] \\ &= \mathrm{Tr}\left[W(B^\dagger \delta B + \delta B^\dagger B)_{\mathrm{Re}} - (iB^\dagger \delta B + i\delta B^\dagger B)_{\mathrm{Re}} \Lambda \right] \\ &= 2\mathrm{Re}\mathrm{Tr}\left[\delta B^\dagger B (W - i\Lambda) \right]. \end{aligned} \tag{S40}$$

Then we let $\partial_B \mathcal{L}(B,\Lambda) = 0$ for arbitrary δB and get $B(W - i\Lambda) = 0$, which further gives

$$W^{\frac{1}{2}} B^\dagger B W^{\frac{1}{2}} = i W^{\frac{1}{2}} B^\dagger B W^{\frac{1}{2}} W^{-\frac{1}{2}} \Lambda W^{-\frac{1}{2}}. \tag{S41}$$

By taking the real part and imaginary part on both sides, we have

$$\begin{aligned} W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Re}} W^{\frac{1}{2}} &= -W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Im}} W^{\frac{1}{2}} W^{-\frac{1}{2}} \Lambda W^{-\frac{1}{2}}, \\ W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Im}} W^{\frac{1}{2}} &= W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Re}} W^{\frac{1}{2}} W^{-\frac{1}{2}} \Lambda W^{-\frac{1}{2}}. \end{aligned} \tag{S42}$$

Note that $W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Re}} W^{\frac{1}{2}}$ is real symmetric, from the first equation we can get $\left[W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Im}} W^{\frac{1}{2}}, W^{-\frac{1}{2}} \Lambda W^{-\frac{1}{2}} \right] = 0$. $W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Re}} W^{\frac{1}{2}}$, $W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Im}} W^{\frac{1}{2}}$ and $W^{-\frac{1}{2}} \Lambda W^{-\frac{1}{2}}$ can thus be simultaneously diagonalized. Let $W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Re}} W^{\frac{1}{2}} = U D_{\mathrm{Re}} U^\dagger$, $W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Im}} W^{\frac{1}{2}} = U D_{\mathrm{Im}} U^\dagger$ and $W^{-\frac{1}{2}} \Lambda W^{-\frac{1}{2}} = U D_\Lambda U^\dagger$, where D_{Re}, D_{Im} and D_Λ are diagonal matrices and U is a unitary matrix. From Eq.(S42), we then have $D_{\mathrm{Re}} = -D_{\mathrm{Im}} D_\Lambda$ and $D_{\mathrm{Im}} = D_{\mathrm{Re}} D_\Lambda$. Taking the absolute values on both sides, we further have $|D_{\mathrm{Re}}| = |D_{\mathrm{Im}}||D_\Lambda|$ and $|D_{\mathrm{Im}}| = |D_{\mathrm{Re}}||D_\Lambda|$. This gives

$$(|D_{\mathrm{Re}}| - |D_{\mathrm{Im}}|)(I + |D_\Lambda|) = 0, \tag{S43}$$

which indicates that $|D_{\mathrm{Re}}| = |D_{\mathrm{Im}}|$. Thus we have

$$\mathrm{Tr}\left[W(B^\dagger B)_{\mathrm{Re}} \right] = \mathrm{Tr}(D_{\mathrm{Re}}) = \mathrm{Tr}|D_{\mathrm{Re}}| = \mathrm{Tr}|D_{\mathrm{Im}}| = \mathrm{Tr}\left| W^{\frac{1}{2}} (B^\dagger B)_{\mathrm{Im}} W^{\frac{1}{2}} \right|. \tag{S44}$$

Since $(R^\dagger R)_{\mathrm{Im}} = 0$, we also have $(B^\dagger B)_{\mathrm{Im}} = -(A^\dagger A)_{\mathrm{Im}}$, thus $\mathrm{Tr}\left[W(B^\dagger B)_{\mathrm{Re}} \right] = \mathrm{Tr}\left| W^{\frac{1}{2}} (A^\dagger A)_{\mathrm{Im}} W^{\frac{1}{2}} \right|$ and

$$\begin{aligned} \mathcal{E}_0' &= \min_R \left\{ \mathrm{Tr}\left[W\left((R^\dagger R)_{\mathrm{Re}} - R^\dagger X - X^\dagger R + X^\dagger X \right) \right] | (R^\dagger R)_{\mathrm{Im}} = 0 \right\} \\ &= \min_{A:|a_j\rangle \in \mathcal{X}, B:|b_j\rangle \in \mathcal{X}^\perp} \left\{ \mathrm{Tr}\left[W\left((A^\dagger A)_{\mathrm{Re}} + (B^\dagger B)_{\mathrm{Re}} - A^\dagger X - X^\dagger A + X^\dagger X \right) \right] | (A^\dagger A)_{\mathrm{Im}} + (B^\dagger B)_{\mathrm{Im}} = 0 \right\} \\ &= \min_{A:|a_j\rangle \in \mathcal{X}} \left\{ \mathrm{Tr}\left[W\left((A^\dagger A)_{\mathrm{Re}} - A^\dagger X - X^\dagger A + X^\dagger X \right) \right] + \mathrm{Tr}\left| W^{\frac{1}{2}} (A^\dagger A)_{\mathrm{Im}} W^{\frac{1}{2}} \right| \right\} \geq \mathcal{E}_0. \end{aligned} \tag{S45}$$

Here the last inequality holds as there exist an additional constraint in $\mathcal{E}_0'$. On the other hand, for $\mathcal{E}_0$, we have

$$\begin{aligned} \mathcal{E}_0 &= \min_{S,R} \left\{ \mathrm{Tr}\left[W\left(S - R^\dagger X - X^\dagger R + X^\dagger X \right) \right] | S \text{ real symmetric}, S \geq R^\dagger R \right\} \\ &= \min_{S,A:|a_j\rangle \in \mathcal{X}, B:|b_j\rangle \in \mathcal{X}^\perp} \left\{ \mathrm{Tr}\left[W\left(S - A^\dagger X - X^\dagger A + X^\dagger X \right) \right] | S \text{ real symmetric}, S \geq A^\dagger A + B^\dagger B \right\} \\ &\geq \min_{S,A:|a_j\rangle \in \mathcal{X}} \left\{ \mathrm{Tr}\left[W\left(S - A^\dagger X - X^\dagger A + X^\dagger X \right) \right] | S \text{ real symmetric}, S \geq A^\dagger A \right\} \\ &= \min_{A:|a_j\rangle \in \mathcal{X}} \left\{ \mathrm{Tr}\left[W\left((A^\dagger A)_{\mathrm{Re}} - A^\dagger X - X^\dagger A + X^\dagger X \right) \right] + \mathrm{Tr}\left| W^{\frac{1}{2}} (A^\dagger A)_{\mathrm{Im}} W^{\frac{1}{2}} \right| \right\} = \mathcal{E}_0', \end{aligned} \tag{S46}$$

where the inequality holds since $S \geq A^\dagger A + B^\dagger B$ indicates that $S \geq A^\dagger A$. Combining with $\mathcal{E}_0' \geq \mathcal{E}_0$, we have $\mathcal{E}_0' = \mathcal{E}_0$. ■

The tightness of $\mathcal{E}_0$ can then be proved by the following lemma.

Lemma 2. *$\mathcal{E}_0$ is tight, i.e., if $(R^\dagger R)_{\mathrm{Im}} = 0$, there always exist a POVM $\{M_m\}$ and a set of $\{f_j(m)\}$, such that $\sum_m f_j(m) M_m|\psi\rangle = |r_j\rangle$ for $R = \begin{pmatrix}|r_1\rangle & |r_2\rangle & \cdots & |r_n\rangle\end{pmatrix}$, and the weighted root-mean-square error*

$$\mathcal{E} = \mathrm{Tr}\left[W\left((R^\dagger R)_{\mathrm{Re}} - R^\dagger X - X^\dagger R + X^\dagger X\right)\right]. \tag{S47}$$

Proof. For arbitrary $R = \begin{pmatrix}|r_1\rangle & |r_2\rangle & \cdots & |r_n\rangle\end{pmatrix}$, by doing Schmidt's orthonormalization on $\{|\psi\rangle, |r_1\rangle, |r_2\rangle, ..., |r_n\rangle\}$, we can obtain an orthonormal basis $\{|u_k\rangle\}_{k=1}^{n_d}$ with $n_d = \min\{n+1, d\}$, and

$$\begin{aligned} &|u_1\rangle = |\psi\rangle, \\ &|u_2\rangle = \mathcal{N}(|r_1\rangle - \langle\psi|r_1\rangle|\psi\rangle), \\ &\quad\vdots \end{aligned} \tag{S48}$$

where $\mathcal{N}$ is a normalization factor which is a real number. Since R_j is Hermitian, we have $\langle\psi|r_j\rangle = \langle\psi|R_j|\psi\rangle \in \mathcal{R}$, and from $(R^\dagger R)_{\mathrm{Im}} = 0$, we also have $\mathrm{Im}\langle r_j|r_k\rangle = 0$, $\forall j,k$. Thus the coefficients in Eq.(S48) are thus all real numbers. Each $|r_j\rangle$ can then be expanded as $|r_j\rangle = \sum_{k=1}^{n_d} \lambda_{jk}|u_k\rangle$with $\lambda_{jk} \in \mathcal{R}$.

We then choose an $n_d \times n_d$ real orthogonal matrix P, with P_{jk} being its jkth element, to obtain a rotated basis $|u_j'\rangle = \sum_{k=1}^{n_d} P_{jk}|u_k\rangle$. By choosing a P such that $P_{j1} \neq 0$, we have $\langle\psi|u_j'\rangle \neq 0$, $\forall j$. In the new basis, $\{|u_j'\rangle\}$, $|r_j\rangle$ can be represented as

$$|r_j\rangle = \sum_{k=1}^{n_d} \lambda_{jk}|u_k\rangle = \sum_{k=1}^{n_d} \lambda_{jk} \sum_{l=1}^{n_d} P_{lk}|u_l'\rangle = \sum_{l=1}^{n_d}\left(\sum_{k=1}^{n_d} \lambda_{jk} P_{lk}\right)|u_l'\rangle. \tag{S49}$$

We can now choose a measurement

$$M_m = |u_m'\rangle\langle u_m'|, \text{ for } 1 \leq m \leq n_d, \quad M_0 = I - \sum_{m=1}^{n_d} |u_m'\rangle\langle u_m'|, \tag{S50}$$

and

$$f_j(m) = \frac{\sum_{k=1}^{n_d} \lambda_{jk} P_{mk}}{\langle u_m'|\psi\rangle} \text{ for } 1 \leq m \leq n_d, \quad f_j(m) = 0 \text{ for } m = 0, \tag{S51}$$

for which we have $\sum_{m=0}^{n_d} f_j(m) M_m|\psi\rangle = \sum_{m=1}^{n_d} f_j(m)\langle u_m'|\psi\rangle|u_m'\rangle + f_j(0)M_0|\psi\rangle = |r_j\rangle$. The bound given by

$$\mathcal{E}_0 = \min_R \left\{\mathrm{Tr}\left[W\left((R^\dagger R)_{\mathrm{Re}} - R^\dagger X - X^\dagger R + X^\dagger X\right)\right] | (R^\dagger R)_{\mathrm{Im}} = 0\right\} \tag{S52}$$

can thus be achieved.

■

A. Tightness for mixed states

In this subsection, we provide an explicit example that for mixed states, $\mathcal{E} \geq \mathcal{E}_0$ is in general not tight.

Take the state ρ and observables X_1, X_2 as

$$\rho = \frac{1}{3}\begin{pmatrix} 1-p & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1+p & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, X_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & -2 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \end{pmatrix}, X_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -\frac{3}{2} \\ 0 & 0 & -\frac{3}{2} & 0 \end{pmatrix}. \tag{S53}$$

Besides the calculation of the SDP $\mathcal{E}_0$, we do a brute-force search over all the possible measurements for the error $\epsilon_1^2 + \epsilon_2^2$ numerically. Then a gap between the SDP bound and a brute-force search can be illustrated as Fig.S1, indicating that for mixed states, $\mathcal{E} \geq \mathcal{E}_0$ is in general not tight.

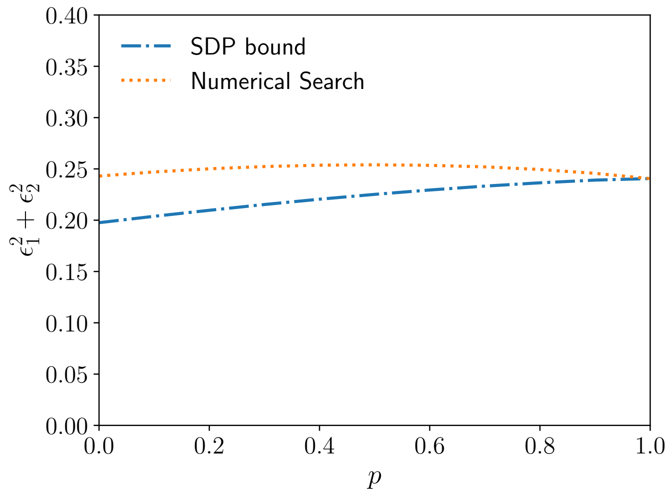

Figure S1. Comparison of the SDP bound and a brute-force search.

S5. ANALYTICAL ERROR-TRADEOFF RELATION AND OPTIMAL MEASUREMENT FOR TWO OBSERVABLES ON PURE STATES

In this section, we analytically solve the optimization in $\mathcal{E}_0$ in the case of two observables with $W = diag(w_1, w_2)$ and $\rho = |\psi\rangle\langle\psi|$ as a pure state, and provide the optimal measurement that saturates the bound.

In the case of two observables, the condition, $(R^\dagger R)_{\mathrm{Im}} = 0$, is equivalent to $\mathrm{Tr}(\rho\,[R_1, R_2]) = 0$ with $\rho = |\psi\rangle\langle\psi|$. The minimization in $\mathcal{E}_0$ can thus be equivalently written as

$$\mathcal{E}_0 = \min_{R_1,R_2}\left\{\sum_{j=1}^{2} w_j \mathrm{Tr}\left[\rho(R_jR_j - R_jX_j - X_jR_j + X_jX_j)\right] | \mathrm{Tr}(\rho\,[R_1,R_2]) = 0, R_1, R_2 \text{ Hermitian}\right\}. \tag{S54}$$

Denote $\boldsymbol{R} = \sqrt{w_1}\tilde{R}_1 + i\sqrt{w_2}\tilde{R}_2$ and $\boldsymbol{X} = \sqrt{w_1}\tilde{X}_1 + i\sqrt{w_2}\tilde{X}_2$, $\mathcal{E}_0$ can then be written as

$$\mathcal{E}_0 = \min_{\boldsymbol{R}}\left\{\mathrm{Tr}(\rho\boldsymbol{R}^\dagger\boldsymbol{R}) - \frac{1}{2}\mathrm{Tr}\left[\rho(\boldsymbol{R}\boldsymbol{X}^\dagger + \boldsymbol{X}^\dagger\boldsymbol{R} + \boldsymbol{R}^\dagger\boldsymbol{X} + \boldsymbol{X}\boldsymbol{R}^\dagger - \boldsymbol{X}\boldsymbol{X}^\dagger - \boldsymbol{X}^\dagger\boldsymbol{X})\right] | \mathrm{Tr}(\rho\boldsymbol{R}^\dagger\boldsymbol{R}) = \mathrm{Tr}(\rho\boldsymbol{R}\boldsymbol{R}^\dagger)\right\}. \tag{S55}$$

The minimization can be solved via the Lagrangian multiplier with

$$\mathcal{L}(\boldsymbol{R},\mu) = \mathrm{Tr}(\rho\boldsymbol{R}^\dagger\boldsymbol{R}) - \frac{1}{2}\mathrm{Tr}\left[\rho(\boldsymbol{R}\boldsymbol{X}^\dagger + \boldsymbol{X}^\dagger\boldsymbol{R} + \boldsymbol{R}^\dagger\boldsymbol{X} + \boldsymbol{X}\boldsymbol{R}^\dagger - \boldsymbol{X}\boldsymbol{X}^\dagger - \boldsymbol{X}^\dagger\boldsymbol{X})\right] + \mu\left[\mathrm{Tr}(\rho\boldsymbol{R}\boldsymbol{R}^\dagger) - \mathrm{Tr}(\rho\boldsymbol{R}^\dagger\boldsymbol{R})\right]. \tag{S56}$$

The matrix derivative of $\mathcal{L}(\boldsymbol{R},\mu)$ with respect to $\boldsymbol{R}$ can be calculated as

$$\begin{aligned}\partial_{\boldsymbol{R}}\mathcal{L}(\boldsymbol{R},\mu) &= \lim_{\varepsilon\to 0}\frac{\mathcal{L}(\boldsymbol{R}+\varepsilon\cdot\delta\boldsymbol{R},\mu) - \mathcal{L}(\boldsymbol{R},\mu)}{\varepsilon} \\ &= \mathrm{Tr}\left\{\delta\boldsymbol{R}\left[\mu\boldsymbol{R}^\dagger\rho + (1-\mu)\rho\boldsymbol{R}^\dagger - \frac{1}{2}\left(\boldsymbol{X}^\dagger\rho + \rho\boldsymbol{X}^\dagger\right)\right] + \delta\boldsymbol{R}^\dagger\left[\mu\rho\boldsymbol{R} + (1-\mu)\boldsymbol{R}\rho - \frac{1}{2}\left(\rho\boldsymbol{X} + \boldsymbol{X}\rho\right)\right]\right\}.\end{aligned} \tag{S57}$$

Then we let $\partial_{\boldsymbol{R}}\mathcal{L}(\boldsymbol{R},\mu) = 0$ for arbitrary $\delta\boldsymbol{R}$ and obtain

$$\mu\rho\boldsymbol{R} + (1-\mu)\boldsymbol{R}\rho - \frac{1}{2}\left(\rho\boldsymbol{X} + \boldsymbol{X}\rho\right) = 0. \tag{S58}$$

By taking trace on both sides of the equation, we have $\langle\psi|\boldsymbol{R}|\psi\rangle = \langle\psi|\boldsymbol{X}|\psi\rangle$. By multiplying $|\psi\rangle$ from either left or right side of Eq.(S58), we get

$$\begin{aligned}(1-\mu)\boldsymbol{R}|\psi\rangle &= \frac{1}{2}(1-2\mu)\langle\psi|\boldsymbol{X}|\psi\rangle|\psi\rangle + \frac{1}{2}\boldsymbol{X}|\psi\rangle, \\ \mu\boldsymbol{R}^\dagger|\psi\rangle &= \frac{1}{2}(2\mu-1)\langle\psi|\boldsymbol{X}^\dagger|\psi\rangle|\psi\rangle + \frac{1}{2}\boldsymbol{X}^\dagger|\psi\rangle.\end{aligned} \tag{S59}$$

Using Eq.(S59) and $\langle\psi|\boldsymbol{R}^\dagger\boldsymbol{R}|\psi\rangle = \langle\psi|\boldsymbol{R}\boldsymbol{R}^\dagger|\psi\rangle$, we can then solve μ. We first consider the cases where $\mu \notin \{0,1\}$. In this case we have

$$\begin{aligned}
\langle\psi|\boldsymbol{R}^\dagger\boldsymbol{R}|\psi\rangle &= \frac{1}{(1-\mu)^2}\left(\frac{1}{2}(1-2\mu)\langle\psi|\boldsymbol{X}^\dagger|\psi\rangle\langle\psi| + \frac{1}{2}\langle\psi|\boldsymbol{X}^\dagger\right)\left(\frac{1}{2}(1-2\mu)\langle\psi|\boldsymbol{X}|\psi\rangle|\psi\rangle + \frac{1}{2}\boldsymbol{X}|\psi\rangle\right)\\
&= \frac{(1-2\mu)(3-2\mu)}{4(1-\mu)^2}|\langle\psi|\boldsymbol{X}|\psi\rangle|^2 + \frac{1}{4(1-\mu)^2}\langle\psi|\boldsymbol{X}^\dagger\boldsymbol{X}|\psi\rangle,\\
\langle\psi|\boldsymbol{R}\boldsymbol{R}^\dagger|\psi\rangle &= \frac{1}{\mu^2}\left(\frac{1}{2}(2\mu-1)\langle\psi|\boldsymbol{X}|\psi\rangle\langle\psi| + \frac{1}{2}\langle\psi|\boldsymbol{X}\right)\left(\frac{1}{2}(2\mu-1)\langle\psi|\boldsymbol{X}^\dagger|\psi\rangle|\psi\rangle + \frac{1}{2}\boldsymbol{X}^\dagger|\psi\rangle\right)\\
&= \frac{(2\mu-1)(2\mu+1)}{4\mu^2}|\langle\psi|\boldsymbol{X}|\psi\rangle|^2 + \frac{1}{4\mu^2}\langle\psi|\boldsymbol{X}\boldsymbol{X}^\dagger|\psi\rangle.
\end{aligned} \tag{S60}$$

$\langle\psi|\boldsymbol{R}^\dagger\boldsymbol{R}|\psi\rangle - \langle\psi|\boldsymbol{R}\boldsymbol{R}^\dagger|\psi\rangle = 0$ then leads to

$$\frac{1}{4}\left(\frac{1}{(1-\mu)^2} - \frac{1}{\mu^2}\right)\alpha + \frac{1}{4}\left(\frac{1}{\mu^2} + \frac{1}{(1-\mu)^2}\right)\beta = 0, \tag{S61}$$

where

$$\begin{aligned}
\alpha &= \frac{1}{2}\langle\psi|(\boldsymbol{X}^\dagger\boldsymbol{X} + \boldsymbol{X}\boldsymbol{X}^\dagger)|\psi\rangle - |\langle\psi|\boldsymbol{X}|\psi\rangle|^2 = w_1(\Delta X_1)^2 + w_2(\Delta X_2)^2,\\
\beta &= \frac{1}{2}\langle\psi|(\boldsymbol{X}^\dagger\boldsymbol{X} - \boldsymbol{X}\boldsymbol{X}^\dagger)|\psi\rangle = i\sqrt{w_1w_2}\langle\psi|[X_1,X_2]|\psi\rangle,
\end{aligned} \tag{S62}$$

here $(\Delta X_j)^2 = \langle\psi|X_j^2|\psi\rangle - \langle\psi|X_j|\psi\rangle^2$ denotes the variance of X_j for $j = 1,2$. From Eq.(S61) we get two solutions,

$$\mu_\pm = \frac{-(\alpha-\beta) \pm \sqrt{\alpha^2-\beta^2}}{2\beta}. \tag{S63}$$

With $\mu = \mu_\pm$, and R_1, R_2 satisfies Eq.(S59) we have

$$\begin{aligned}
\langle\psi|\boldsymbol{R}\boldsymbol{X}^\dagger|\psi\rangle &= \frac{1}{\mu}\left(\frac{1}{2}(2\mu-1)\langle\psi|\boldsymbol{X}|\psi\rangle\langle\psi| + \frac{1}{2}\langle\psi|\boldsymbol{X}\right)\boldsymbol{X}^\dagger|\psi\rangle = \frac{2\mu-1}{2\mu}|\langle\psi|\boldsymbol{X}|\psi\rangle|^2 + \frac{1}{2\mu}\langle\psi|\boldsymbol{X}\boldsymbol{X}^\dagger|\psi\rangle,\\
\langle\psi|\boldsymbol{X}^\dagger\boldsymbol{R}|\psi\rangle &= \frac{1}{1-\mu}\langle\psi|\boldsymbol{X}^\dagger\left(\frac{1}{2}(1-2\mu)\langle\psi|\boldsymbol{X}|\psi\rangle|\psi\rangle + \frac{1}{2}\boldsymbol{X}|\psi\rangle\right) = \frac{1-2\mu}{2(1-\mu)}|\langle\psi|\boldsymbol{X}|\psi\rangle|^2 + \frac{1}{2(1-\mu)}\langle\psi|\boldsymbol{X}^\dagger\boldsymbol{X}|\psi\rangle,
\end{aligned} \tag{S64}$$

and

$$\begin{aligned}
\mathcal{E}_0 =&\langle\psi|\boldsymbol{R}^\dagger\boldsymbol{R}|\psi\rangle - \frac{1}{2}\langle\psi|(\boldsymbol{R}\boldsymbol{X}^\dagger + \boldsymbol{X}^\dagger\boldsymbol{R} + \boldsymbol{R}^\dagger\boldsymbol{X} + \boldsymbol{X}\boldsymbol{R}^\dagger - \boldsymbol{X}\boldsymbol{X}^\dagger - \boldsymbol{X}^\dagger\boldsymbol{X})|\psi\rangle\\
=&\left(\frac{(1-2\mu)(3-2\mu)}{4(1-\mu)^2}|\langle\psi|\boldsymbol{X}|\psi\rangle|^2 + \frac{1}{4(1-\mu)^2}\langle\psi|\boldsymbol{X}^\dagger\boldsymbol{X}|\psi\rangle\right) + \frac{1}{2}\langle\psi|(\boldsymbol{X}\boldsymbol{X}^\dagger + \boldsymbol{X}^\dagger\boldsymbol{X})|\psi\rangle\\
&-\left(\frac{2\mu-1}{2\mu}|\langle\psi|\boldsymbol{X}|\psi\rangle|^2 + \frac{1}{2\mu}\langle\psi|\boldsymbol{X}\boldsymbol{X}^\dagger|\psi\rangle\right) - \left(\frac{1-2\mu}{2(1-\mu)}|\langle\psi|\boldsymbol{X}|\psi\rangle|^2 + \frac{1}{2(1-\mu)}\langle\psi|\boldsymbol{X}^\dagger\boldsymbol{X}|\psi\rangle\right)\\
=&\left(\frac{(1-2\mu)(3-2\mu)}{4(1-\mu)^2} - \frac{2\mu-1}{2\mu} - \frac{1-2\mu}{2(1-\mu)}\right)|\langle\psi|\boldsymbol{X}|\psi\rangle|^2 + \frac{1}{2}\langle\psi|(\boldsymbol{X}\boldsymbol{X}^\dagger + \boldsymbol{X}^\dagger\boldsymbol{X})|\psi\rangle\\
&+\left(\frac{1}{4(1-\mu)^2} - \frac{1}{2(1-\mu)}\right)\langle\psi|\boldsymbol{X}^\dagger\boldsymbol{X}|\psi\rangle - \frac{1}{2\mu}\langle\psi|\boldsymbol{X}\boldsymbol{X}^\dagger|\psi\rangle\\
=&\left(1 - \frac{2-3\mu}{4\mu(1-\mu)^2}\right)\alpha + \frac{4\mu^2-5\mu+2}{4\mu(1-\mu)^2}\beta\\
=&\frac{1}{2}\left(\alpha \mp \sqrt{\alpha^2-\beta^2}\right).
\end{aligned} \tag{S65}$$

The minimal $\mathcal{E}_0 = \frac{1}{2}\left(\alpha - \sqrt{\alpha^2-\beta^2}\right)$ is achieved with $\mu = \mu_+$.

In the cases where $\mu \in \{0,1\}$, Eq.(S59) have solutions only if $\alpha = \beta$. For example, if $\mu = 0$, Eq.(S59) becomes

$$\begin{aligned}
\boldsymbol{R}|\psi\rangle &= \frac{1}{2}\langle\psi|\boldsymbol{X}|\psi\rangle|\psi\rangle + \frac{1}{2}\boldsymbol{X}|\psi\rangle,\\
0 &= -\frac{1}{2}\langle\psi|\boldsymbol{X}^\dagger|\psi\rangle|\psi\rangle + \frac{1}{2}\boldsymbol{X}^\dagger|\psi\rangle.
\end{aligned} \tag{S66}$$

Multiply $\langle\psi|\boldsymbol{X}$ from the left of the second equation, we have $\langle\psi|\boldsymbol{X}\boldsymbol{X}^\dagger|\psi\rangle = |\langle\psi|\boldsymbol{X}|\psi\rangle|^2$, which gives $\alpha=\beta$. In this case

$$\begin{aligned}\langle\psi|\boldsymbol{R}^\dagger\boldsymbol{R}|\psi\rangle &= \frac{1}{4}(3|\langle\psi|\boldsymbol{X}|\psi\rangle|^2 + \langle\psi|\boldsymbol{X}^\dagger\boldsymbol{X}|\psi\rangle),\\ \langle\psi|\boldsymbol{R}\boldsymbol{X}^\dagger|\psi\rangle &= |\langle\psi|\boldsymbol{X}|\psi\rangle|^2,\\ \langle\psi|\boldsymbol{R}^\dagger\boldsymbol{X}|\psi\rangle &= \frac{1}{2}(\langle\psi|\boldsymbol{X}^\dagger\boldsymbol{X}|\psi\rangle + |\langle\psi|\boldsymbol{X}|\psi\rangle|^2).\end{aligned} \tag{S67}$$

Then

$$\begin{aligned}\mathcal{E}_0 =&\langle\psi|\boldsymbol{R}^\dagger\boldsymbol{R}|\psi\rangle - \frac{1}{2}\langle\psi|(\boldsymbol{R}\boldsymbol{X}^\dagger + \boldsymbol{X}^\dagger\boldsymbol{R} + \boldsymbol{R}^\dagger\boldsymbol{X} + \boldsymbol{X}\boldsymbol{R}^\dagger - \boldsymbol{X}\boldsymbol{X}^\dagger - \boldsymbol{X}^\dagger\boldsymbol{X})|\psi\rangle\\ =&\frac{1}{4}(\langle\psi|\boldsymbol{X}^\dagger\boldsymbol{X}|\psi\rangle - |\langle\psi|\boldsymbol{X}|\psi\rangle|^2)\\ =&\frac{1}{4}(\langle\psi|\boldsymbol{X}^\dagger\boldsymbol{X}|\psi\rangle - \langle\psi|\boldsymbol{X}\boldsymbol{X}^\dagger|\psi\rangle)\\ =&\frac{1}{2}\alpha.\end{aligned} \tag{S68}$$

Similarly for $\mu=1$, we also have $\alpha=\beta$ and $\mathcal{E}_0=\frac{1}{2}\alpha$. Thus $\mu\in\{0,1\}$ can only happen when $\alpha=\beta$, in which case the minimal $\mathcal{E}_0$ can also be written as $\mathcal{E}_0 = \frac{1}{2}(\alpha-\sqrt{\alpha^2-\beta^2})$. To summarize, we have the tight bound

$$w_1\epsilon_1^2 + w_2\epsilon_2^2 \geq \frac{1}{2}\left(\alpha-\sqrt{\alpha^2-\beta^2}\right), \tag{S69}$$

where $\alpha = w_1(\Delta X_1)^2 + w_2(\Delta X_2)^2,\ \beta = i\sqrt{w_1w_2}\langle\psi|[X_1,X_2]|\psi\rangle$.

We note that since Ozawa's relation is tight for two observables on pure states, the tight bound can also be obtained from the Ozawa's relation, which is presented in the next section. The advantage of the derivation here is it can directly use Lemma 2 in Appendix S4 to construct the optimal measurement. Without loss of generality, we can assume $\langle\psi|X_1|\psi\rangle = \langle\psi|X_2|\psi\rangle = 0$, which is equivalent to replacing $X_j = X_j - \langle\psi|X_j|\psi\rangle I$ for $j=1,2$ and does not affect the root-mean-square error and the optimal measurement. Then with Eq.(S59), we have

$$\begin{aligned}(1-\mu)(|r_1\rangle + i|r_2\rangle) &= \frac{1}{2}(|x_1\rangle + i|x_2\rangle),\\ \mu(|r_1\rangle - i|r_2\rangle) &= \frac{1}{2}(|x_1\rangle - i|x_2\rangle),\end{aligned} \tag{S70}$$

which gives $|r_1\rangle$, $|r_2\rangle$ as

$$\begin{aligned}|r_1\rangle &= \frac{1}{4(1-\mu)}(|x_1\rangle + i|x_2\rangle) + \frac{1}{4\mu}(|x_1\rangle - i|x_2\rangle) = \frac{1}{4\mu(1-\mu)}|x_1\rangle - i\frac{1-2\mu}{4\mu(1-\mu)}|x_2\rangle,\\ |r_2\rangle &= \frac{1}{i4(1-\mu)}(|x_1\rangle + i|x_2\rangle) - \frac{1}{i4\mu}(|x_1\rangle - i|x_2\rangle) = i\frac{1-2\mu}{4\mu(1-\mu)}|x_1\rangle + \frac{1}{4\mu(1-\mu)}|x_2\rangle,\end{aligned} \tag{S71}$$

with $|x_j\rangle = X_j|\psi\rangle$, $j=1,2$, $\mu=\mu_+=\frac{-(\alpha-\beta)+\sqrt{\alpha^2-\beta^2}}{2\beta}$. By finding the orthonormal basis $|u'_m\rangle$ of the span $\{|\psi\rangle, |r_1\rangle, |r_2\rangle\}$ such that $\langle\psi|u'_m\rangle \neq 0$, $\forall m$, the optimal measurement $\{M_m\}$ and optimal $\{f_j(m)\}$ can be obtained as Eq.(S50) and Eq.(S51).

S6. TIGHTER ANALYTICAL BOUND FOR TWO OBSERVABLES ON MIXED STATES

In this section, we derive a tighter analytical bound for two observables on mixed states than the Ozawa's bound.

A. Minimal weighted error from Ozawa's relation

For the comparison, we first identify the minimal weighted error, $w_1\epsilon_1^2 + w_2\epsilon_2^2$, derived from the Ozawa's relation,

$$\epsilon_1^2\cdot(\Delta X_2)^2 + (\Delta X_1)^2\cdot\epsilon_2^2 + 2\sqrt{(\Delta X_1)^2\cdot(\Delta X_2)^2 - c_{12}^2}\,\epsilon_1\cdot\epsilon_2 \geq c_{12}^2, \tag{S72}$$

where $c_{12}=\frac{1}{2}\|\sqrt{\rho}[X_1,X_2]\sqrt{\rho}\|_1$.

The minimal weighted error can be obtained via the Lagrange multiplier with

$$\mathcal{L}(\epsilon_1,\epsilon_2,\lambda)=w_1\epsilon_1^2+w_2\epsilon_2^2+\lambda\left(\epsilon_1^2\cdot(\Delta X_2)^2+(\Delta X_1)^2\cdot\epsilon_2^2+2\sqrt{(\Delta X_1)^2\cdot(\Delta X_2)^2-c_{12}^2}\epsilon_1\cdot\epsilon_2-c_{12}^2\right), \tag{S73}$$

which gives

$$\begin{aligned}\partial_{\epsilon_1}\mathcal{L}(\epsilon_1,\epsilon_2,\lambda)&=2w_1\epsilon_1+\lambda\left(2\epsilon_1\cdot(\Delta X_2)^2+2\sqrt{(\Delta X_1)^2\cdot(\Delta X_2)^2-c_{12}^2}\epsilon_2\right)=0,\\ \partial_{\epsilon_2}\mathcal{L}(\epsilon_1,\epsilon_2,\lambda)&=2w_2\epsilon_2+\lambda\left(2\epsilon_2\cdot(\Delta X_1)^2+2\sqrt{(\Delta X_1)^2\cdot(\Delta X_2)^2-c_{12}^2}\epsilon_1\right)=0,\end{aligned} \tag{S74}$$

and $\lambda(\epsilon_1^2\cdot(\Delta X_2)^2+(\Delta X_1)^2\cdot\epsilon_2^2+2\sqrt{(\Delta X_1)^2\cdot(\Delta X_2)^2-c_{12}^2}\epsilon_1\cdot\epsilon_2-c_{12}^2)=0$. If $\lambda=0$, we have $\epsilon_1=\epsilon_2=0$, which can only satisfy the inequality in Eq.(S72) when $c_{12}=0$. If $\lambda\neq 0$, we then have $\epsilon_1^2\cdot(\Delta X_2)^2+(\Delta X_1)^2\cdot\epsilon_2^2+2\sqrt{(\Delta X_1)^2\cdot(\Delta X_2)^2-c_{12}^2}\epsilon_1\cdot\epsilon_2-c_{12}^2=0$, which, combines with Eq.(S74), gives the optimal ϵ_1 and ϵ_2 as

$$\begin{aligned}\epsilon_1^\star=&\sqrt{\frac{1}{2}(\Delta X_1)^2-\frac{1}{2\sqrt{\alpha^2-\beta^2}}(w_1(\Delta X_1)^4+w_2(\Delta X_1)^2\cdot(\Delta X_2)^2-2w_2c_{12}^2)},\\ \epsilon_2^\star=&\sqrt{\frac{1}{2}(\Delta X_2)^2-\frac{1}{2\sqrt{\alpha^2-\beta^2}}(w_2(\Delta X_2)^4+w_1(\Delta X_1)^2\cdot(\Delta X_2)^2-2w_1c_{12}^2)},\end{aligned} \tag{S75}$$

where $\alpha=w_1(\Delta X_1)^2+w_2(\Delta X_2)^2$, $\beta=2\sqrt{w_1w_2}c_{12}$. Note that when $c_{12}=0$, $\epsilon_1^\star=\epsilon_2^\star=0$ are also optimal. Eq.(S75) thus give the optimal value for all cases. By substituting $\epsilon_1^\star$ and $\epsilon_2^\star$ in $w_1\epsilon_1^2+w_2\epsilon_2^2$, we then obtain the bound on the minimal weighted error from the Ozawa's relation as

$$w_1\epsilon_1^2+w_2\epsilon_2^2\geq\mathcal{E}_{Ozawa}\equiv w_1\epsilon_1^{\star 2}+w_2\epsilon_2^{\star 2}=\frac{1}{2}\left(\alpha-\sqrt{\alpha^2-\beta^2}\right), \tag{S76}$$

where $\alpha=w_1(\Delta X_1)^2+w_2(\Delta X_2)^2$, $\beta=\sqrt{w_1w_2}\|\sqrt{\rho}[X_1,X_2]\sqrt{\rho}\|_1$. Note that $\epsilon_1^\star$ and $\epsilon_2^\star$ saturates the Ozawa's relation with $\epsilon_1^{\star 2}\cdot(\Delta X_2)^2+(\Delta X_1)^2\cdot\epsilon_2^{\star 2}+2\sqrt{(\Delta X_1)^2\cdot(\Delta X_2)^2-c_{12}^2}\epsilon_1^\star\cdot\epsilon_2^\star=c_{12}^2$, the bound in Eq.(S76) is thus as tight as the Ozawa's relation. It is saturable for pure state, however, for mixed states, however, the bound is not tight in general since the Ozawa's relation is in general not tight for mixed states.

B. Tighter analytical bound for the weighted error

We now derive a tighter analytical bound. Given any mixed state ρ, the minimal $\mathcal{E}$ is lower bounded by $\mathcal{E}_0$ in Eq.(S27). Given any $\{|u_q\rangle\}$ that satistifies $\sum_q|u_q\rangle\langle u_q|=I$, we have $\rho=\sum_q\sqrt{\rho}|u_q\rangle\langle u_q|\sqrt{\rho}=\sum_q\lambda_q|\phi_q\rangle\langle\phi_q|$, here

$$\lambda_q=\langle u_q|\rho|u_q\rangle,\quad |\phi_q\rangle=\frac{\sqrt{\rho}|u_q\rangle}{\sqrt{\langle u_q|\rho|u_q\rangle}}. \tag{S77}$$

Then we have $\mathcal{E}_0\geq\sum_q\lambda_q\mathcal{E}_{|\phi_q\rangle}$, with

$$\begin{aligned}\mathcal{E}_{|\phi_q\rangle}=&\min_{\mathbb{S},\{R_j\}_{j=1}^n}\ \mathrm{Tr}\left[(W\otimes|\phi_q\rangle\langle\phi_q|)(\mathbb{S}-\mathbb{R}\mathbb{X}^\dagger-\mathbb{X}\mathbb{R}^\dagger+\mathbb{X}\mathbb{X}^\dagger)\right]\\ &\text{subject to}\quad \mathbb{S}_{jk}=\mathbb{S}_{kj}=\mathbb{S}_{jk}^\dagger,\ \forall j,k\\ &\qquad\qquad\quad R_j=R_j^\dagger,\ \forall j\\ &\qquad\qquad\quad \begin{pmatrix}I & \mathbb{R}^\dagger\\ \mathbb{R} & \mathbb{S}\end{pmatrix}\geq 0.\end{aligned} \tag{S78}$$

For two observables this can be analytically solved as in previous section, which gives

$$\mathcal{E}_{|\phi_q\rangle}=\frac{1}{2}\left(\alpha_q-\sqrt{\alpha_q^2-\beta_q^2}\right), \tag{S79}$$

where $\alpha_q = w_1(\Delta_{|\phi_q\rangle} X_1)^2 + w_2(\Delta_{|\phi_q\rangle} X_2)^2$, $\beta_q = i\sqrt{w_1 w_2}\langle\phi_q|[X_1, X_2]|\phi_q\rangle$, $(\Delta_{|\phi_q\rangle} X_j)^2 = \langle\phi_q|X_j^2|\phi_q\rangle - \langle\phi_q|X_j|\phi_q\rangle^2$. Since $\mathcal{E}_0 \geq \sum_q \lambda_q \mathcal{E}_{|\phi_q\rangle}$, we then get an analytical bound $\mathcal{E} \geq \mathcal{E}_0 \geq \mathcal{E}_A$ with

$$\mathcal{E}_A = \sum_q \lambda_q \mathcal{E}_{|\phi_q\rangle} = \sum_q \frac{1}{2}\lambda_q \left(\alpha_q - \sqrt{\alpha_q^2 - \beta_q^2}\right). \tag{S80}$$

Any choice of $\{|u_q\rangle\}$ with $\sum_q |u_q\rangle\langle u_q| = I$ can be used to obtain an analytical bound. Specifically by choosing $|u_q\rangle$ as the eigenvectors of $\sqrt{\rho}[X_1, X_2]\sqrt{\rho}$, we can get an analytical bound that is in general tighter than Ozawa's bound for mixed states.

For the comparison of $\mathcal{E}_A$ with $\mathcal{E}_{Ozawa}$, we first note that

$$\sum_q \lambda_q|\beta_q| = \sum_q \langle u_q|\rho|u_q\rangle \left|\sqrt{w_1 w_2}\frac{\langle u_q|\sqrt{\rho}[X_1, X_2]\sqrt{\rho}|u_q\rangle}{\langle u_q|\rho|u_q\rangle}\right| = \sqrt{w_1 w_2}\|\sqrt{\rho}[X_1, X_2]\sqrt{\rho}\|_1 = \beta, \tag{S81}$$

$$\begin{aligned}
\sum_q \lambda_q \alpha_q &= \sum_{j=1,2} w_j \sum_q \langle u_q|\rho|u_q\rangle (\Delta_{|\phi_q\rangle} X_j)^2 \\
&= \sum_{j=1,2} w_j \sum_q \langle u_q|\rho|u_q\rangle \left(\frac{\langle u_q|\sqrt{\rho}X_j^2\sqrt{\rho}|u_q\rangle}{\langle u_q|\rho|u_q\rangle} - \langle\phi_q|X_j|\phi_q\rangle^2\right) \\
&\leq \sum_{j=1,2} w_j \sum_q \langle u_q|\rho|u_q\rangle \left(\frac{\langle u_q|\sqrt{\rho}X_j^2\sqrt{\rho}|u_q\rangle}{\langle u_q|\rho|u_q\rangle}\right) = \sum_{j=1,2} w_j \mathrm{Tr}(\rho X_j^2) = \sum_{j=1,2} w_j (\Delta X_j)^2,
\end{aligned} \tag{S82}$$

where without loss of generality we assumed that $\mathrm{Tr}(\rho X_1) = \mathrm{Tr}(\rho X_2) = 0$. We thus have $\sum_q \lambda_q \alpha_q \leq \alpha$ and $\sum_q \lambda_q |\beta_q| = \beta$. We also have (note $\alpha_j \geq |\beta_j|$)

$$\left(\alpha_1 - \sqrt{\alpha_1^2 - \beta_1^2}\right) + \left(\alpha_2 - \sqrt{\alpha_2^2 - \beta_2^2}\right) \geq \left(\alpha_1 + \alpha_2 - \sqrt{(\alpha_1 + \alpha_2)^2 - (|\beta_1| + |\beta_2|)^2}\right), \tag{S83}$$

which can be directly verified as

$$\begin{aligned}
&\Leftrightarrow \quad \sqrt{\alpha_1^2 - \beta_1^2} + \sqrt{\alpha_2^2 - \beta_2^2} \leq \sqrt{(\alpha_1 + \alpha_2)^2 - (|\beta_1| + |\beta_2|)^2}, \\
&\Leftrightarrow \quad \alpha_1^2 - \beta_1^2 + \alpha_2^2 - \beta_2^2 + 2\sqrt{(\alpha_1^2 - \beta_1^2)(\alpha_2^2 - \beta_2^2)} \leq \alpha_1^2 - \beta_1^2 + \alpha_2^2 - \beta_2^2 + 2(\alpha_1\alpha_2 - |\beta_1\beta_2|), \\
&\Leftrightarrow \quad (\alpha_1^2 - \beta_1^2)(\alpha_2^2 - \beta_2^2) \leq (\alpha_1\alpha_2 - |\beta_1\beta_2|)^2, \\
&\Leftrightarrow \quad \alpha_1^2\beta_2^2 + \beta_1^2\alpha_2^2 \geq 2\alpha_1\alpha_2|\beta_1\beta_2| \\
&\Leftrightarrow \quad (\alpha_1|\beta_2| - |\beta_1|\alpha_2)^2 \geq 0.
\end{aligned} \tag{S84}$$

We thus have

$$\begin{aligned}
\mathcal{E}_A &= \frac{1}{2}\sum_q \lambda_q \left(\alpha_q - \sqrt{\alpha_q^2 - \beta_q^2}\right) = \frac{1}{2}\sum_q \left((\lambda_q\alpha_q) - \sqrt{(\lambda_q\alpha_q)^2 - (\lambda_q|\beta_q|)^2}\right) \\
&\geq \frac{1}{2}\left((\sum_q \lambda_q\alpha_q) - \sqrt{(\sum_q \lambda_q\alpha_q)^2 - (\sum_q \lambda_q|\beta_q|)^2}\right) \\
&\geq \frac{1}{2}\left(\alpha - \sqrt{\alpha^2 - \beta^2}\right) = \mathcal{E}_{Ozawa},
\end{aligned} \tag{S85}$$

where the last inequality we used the fact that $f(\alpha) = \alpha - \sqrt{\alpha^2 - \beta^2}$, $\alpha \in (\beta, \infty)$, is a monotonically non-increasing function since $f'(\alpha) = 1 - \frac{\alpha}{\sqrt{\alpha^2 - \beta^2}} \leq 0$. The analytical bound, $\mathcal{E}_A = \sum_q \frac{1}{2}\lambda_q \left(\alpha_q - \sqrt{\alpha_q^2 - \beta_q^2}\right)$, is thus always tighter than the bound obtained from the Ozawa's relation, $\mathcal{E}_{Ozawa} = \frac{1}{2}\left(\alpha - \sqrt{\alpha^2 - \beta^2}\right)$.

C. Example

Here we provide an example to illustrate the difference among the SDP bound $\mathcal{E}_0$ (Eq.(S27)), the analytical bound $\mathcal{E}_A$, and the bound obtained from the Ozawa's relation, $\mathcal{E}_{Ozawa}$.

Take ρ and three observables, X_1, X_2, X_3 as

$$\begin{aligned} \rho &= \begin{pmatrix} \frac{p}{2} & 0 & 0 \\ 0 & 1-p & 0 \\ 0 & 0 & \frac{p}{2} \end{pmatrix}, \ X_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\ X_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \ X_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \end{aligned} \tag{S86}$$

Here X_1, X_2, X_3 are the spin-1 operators with $[X_j, X_k] \neq 0$ for all $j \neq k$. In this case we have $(\Delta X_1)^2 = (\Delta X_2)^2 = 2-p$, $(\Delta X_3)^2 = p$, and

$$\begin{aligned} c_{12} &= \frac{1}{2}\|\sqrt{\rho}[X_1, X_2]\sqrt{\rho}\|_1 = p, \\ c_{23} &= \frac{1}{2}\|\sqrt{\rho}[X_2, X_3]\sqrt{\rho}\|_1 = \sqrt{p(1-p)}, \\ c_{31} &= \frac{1}{2}\|\sqrt{\rho}[X_3, X_1]\sqrt{\rho}\|_1 = \sqrt{p(1-p)}. \end{aligned} \tag{S87}$$

For the simultaneous measurement of (X_1, X_2), (X_2, X_3), (X_3, X_1), the Ozawa's relation gives

$$\begin{aligned} \epsilon_1^2 + \epsilon_2^2 &\geq \mathcal{E}_{Ozawa}(X_1, X_2) = 2 - p - 2\sqrt{1-p}, \\ \epsilon_2^2 + \epsilon_3^2 &\geq \mathcal{E}_{Ozawa}(X_2, X_3) = 1 - \sqrt{1-p(1-p)}, \\ \epsilon_3^2 + \epsilon_1^2 &\geq \mathcal{E}_{Ozawa}(X_3, X_1) = 1 - \sqrt{1-p(1-p)}. \end{aligned} \tag{S88}$$

While $\mathcal{E}_A$ can be calculated by substituting $\{|u_q\rangle\}$ with the eigenvectors of $\sqrt{\rho}[X_j, X_k]\sqrt{\rho}$. Specifically for (X_2, X_3), we have $|u_0\rangle = (1, \sqrt{2}, 1)^T/2$, $|u_1\rangle = (1, -\sqrt{2}, 1)^T/2$, $|u_2\rangle = (1, 0, -1)^T/\sqrt{2}$, which gives

$$\begin{aligned} \epsilon_2^2 + \epsilon_3^2 \geq \mathcal{E}_A(X_2, X_3) &= \sum_{q=0}^{2} \frac{1}{2}\lambda_q \left(\alpha_q - \sqrt{\alpha_q^2 - \beta_q^2}\right) \\ &= \frac{2-p}{4}\left(\frac{4-3p}{2-p} - \sqrt{\left(\frac{4-3p}{2-p}\right)^2 - \left(\frac{4\sqrt{p(1-p)}}{2-p}\right)^2}\right) \\ &= \frac{1}{4}(4 - 3p - |4-5p|). \end{aligned} \tag{S89}$$

Similarly for (X_1, X_2) and (X_3, X_1), we have

$$\begin{aligned} \epsilon_1^2 + \epsilon_2^2 &\geq \mathcal{E}_A(X_1, X_2) = p, \\ \epsilon_3^2 + \epsilon_1^2 &\geq \mathcal{E}_A(X_3, X_1) = \frac{1}{4}(4 - 3p - |4-5p|). \end{aligned} \tag{S90}$$

In Fig. S2, we illustrate the comparisons among the three bounds: (i) $\mathcal{E}_0$ that can be computed with SDP; (ii) $\mathcal{E}_A$ that can be analytically obtained and (iii) $\mathcal{E}_{Ozawa}$ that obtained from the Ozawa's relation. It can be seen that for the simultaneous measurement of two observables, we already have $\mathcal{E}_0 > \mathcal{E}_A > \mathcal{E}_{Ozawa}$. While for the simultaneous measurement of three observables, the SDP bound is tighter than the bound obtained by simply adding the results for each pairs of the observables.

S7. COMPARISON OF THE ANALYTICAL BOUND FOR MORE THAN TWO OBSERVABLES WITH THE OZAWA'S BOUND ON PURE STATES

We have shown that in the case of two observables, $\mathcal{E}_A = \mathcal{E}_{Ozawa}$ for pure states and $\mathcal{E}_A \geq \mathcal{E}_{Ozawa}$ for mixed states. Thus for mixed states we already get tighter analytical bound than the Ozawa's bound. In this section we show that

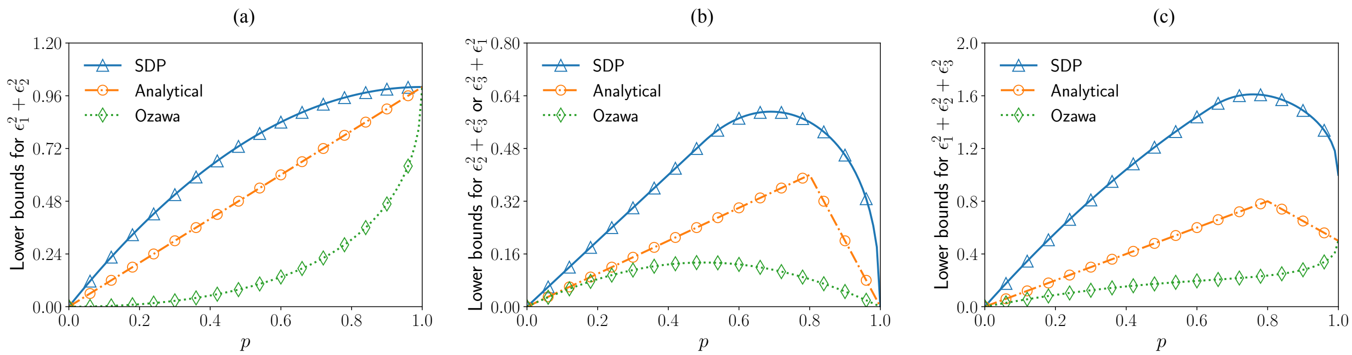

Figure S2. Lower bounds for the errors of the simultaneous measurement of (a) (X_1, X_2), (b) (X_2, X_3) or (X_3, X_1), (c) (X_1, X_2, X_3). The bounds for (X_2, X_3) and (X_3, X_1) are the same, so they are plotted in the same figure (b). The analytical and Ozawa's bounds for (X_1, X_2, X_3) are obtained by summing the bounds for all pairs (X_j, X_k), while the SDP bound can be directly obtained via the SDP for three observables.

for pure states the analytical bound obtained for an arbitrary number of observables,

$$\mathrm{Tr}(S_{\mathrm{Re}}^{-1} Q_{\mathrm{Re}}) \geq \left(\sqrt{\| S_{\mathrm{Re}}^{-\frac{1}{2}} \tilde{S}_{\mathrm{Im}} S_{\mathrm{Re}}^{-\frac{1}{2}} \|_F + 1} - 1 \right)^2, \tag{S91}$$

is always tighter than the simple summation of the Ozawa's bound when the number of observables is bigger than 4 $(n \geq 5)$.

Since $\mathrm{Tr}(S_{\mathrm{Re}}^{-1} Q_{\mathrm{Re}})$ does not change under linear transformations, without loss of generality, we can assume $S_{\mathrm{Re}} = I$, the bound then becomes

$$\sum_j \epsilon_j^2 \geq \left(\sqrt{\| \tilde{S}_{\mathrm{Im}} \|_F + 1} - 1 \right)^2. \tag{S92}$$

For pure state $\rho = |\psi\rangle\langle\psi|$, with arbitrary $\{|u_q\rangle\}$ we have $(S_{u_q})_{jk} = \langle\psi|X_j X_k|\psi\rangle |\langle\psi|u_q\rangle|^2$, and $(S_{u_q,\mathrm{Im}})_{jk} = \frac{1}{2i}\langle\psi|[X_j, X_k]|\psi\rangle |\langle\psi|u_q\rangle|^2$. The matrices $S_{u_q,\mathrm{Im}}$ for different q then equal to a common antisymmetric matrix multiplied by non-negative numbers $|\langle\psi|u_q\rangle|^2$. In this case the optimal choice is not to take any transpose. With arbitrary $\{|u_q\rangle\}$ with $\sum_q |u_q\rangle\langle u_q| = I$, we have $\sum_q |\langle\psi|u_q\rangle|^2 = 1$, and

$$(\tilde{S}_{\mathrm{Im}})_{jk} = \sum_q (S_{u_q,\mathrm{Im}})_{jk} = \sum_q \frac{1}{2i} \langle\psi|[X_j, X_k]|\psi\rangle |\langle\psi|u_q\rangle|^2 = \frac{1}{2i}\langle\psi|[X_j, X_k]|\psi\rangle \tag{S93}$$

From the Ozawa's relation, we show in Appendix S6 B that

$$w_1 \epsilon_j^2 + w_2 \epsilon_k^2 \geq \mathcal{E}_{Ozawa} \equiv \frac{1}{2}\left(\alpha - \sqrt{\alpha^2 - \beta^2} \right), \tag{S94}$$

where $\alpha = w_1(\Delta X_j)^2 + w_2(\Delta X_k)^2$, $\beta = \sqrt{w_1 w_2} \| \sqrt{\rho}[X_j, X_k]\sqrt{\rho} \|_1$. For the comparison we choose $w_1 = w_2 = 1$ and for normalized observables with $S_{\mathrm{Re}} = I$ we have $(\Delta X_j)^2 = 1$, $\forall j$. For pure state, $\rho = |\psi\rangle\langle\psi|$, the Ozawa's bound is then given by

$$\epsilon_j^2 + \epsilon_k^2 \geq \mathcal{E}_{Ozawa}(X_j, X_k) \equiv \frac{1}{2}\left(2 - \sqrt{4 - |\langle\psi|[X_j, X_k]|\psi\rangle|^2} \right) = 1 - \sqrt{1 - \frac{1}{2}|\langle\psi|[X_j, X_k]|\psi\rangle|^2}, \tag{S95}$$

By summing all pairs, we obtain

$$\sum_l \epsilon_l^2 \geq \frac{1}{n-1} \sum_{j,k} \left(1 - \sqrt{1 - \frac{1}{2}|\langle\psi|[X_j, X_k]|\psi\rangle|^2} \right). \tag{S96}$$

Since $1 - \frac{1}{2}|\langle\psi|[X_j, X_k]|\psi\rangle|^2 \leq 1$, we have $\sqrt{1 - \frac{1}{2}|\langle\psi|[X_j, X_k]|\psi\rangle|^2} \geq 1 - \frac{1}{2}|\langle\psi|[X_j, X_k]|\psi\rangle|^2$ and

$$\sum_{j,k} \left(1 - \sqrt{1 - \frac{1}{2}|\langle\psi|[X_j, X_k]|\psi\rangle|^2} \right) \leq \sum_{j,k} \frac{1}{2}|\langle\psi|[X_j, X_k]|\psi\rangle|^2 = \frac{1}{2}\|\tilde{S}_{\mathrm{Im}}\|_F^2. \tag{S97}$$

The bound in Eq.(S92) is then tighter than the bound in Eq.(S96) when

$$\left(\sqrt{\|\tilde{S}_{\rm Im}\|_F+1}-1\right)^2 \geq \frac{1}{2(n-1)}\|\tilde{S}_{\rm Im}\|_F^2 \geq \frac{1}{n-1}\sum_{j,k}(1-\sqrt{1-\frac{1}{2}|\langle\psi|[X_j,X_k]|\psi\rangle|^2}). \tag{S98}$$

The first inequality is equivalent to

$$\begin{aligned}
&\sqrt{\|\tilde{S}_{\rm Im}\|_F+1}-1 \geq \frac{1}{\sqrt{2(n-1)}}\|\tilde{S}_{\rm Im}\|_F \\
\Leftrightarrow\quad & \frac{1}{\sqrt{2(n-1)}}(\|\tilde{S}_{\rm Im}\|_F+1)-\frac{1}{\sqrt{2(n-1)}}-\sqrt{\|\tilde{S}_{\rm Im}\|_F+1}+1\leq 0 \\
\Leftrightarrow\quad & (\|\tilde{S}_{\rm Im}\|_F+1)-\sqrt{2(n-1)}\sqrt{\|\tilde{S}_{\rm Im}\|_F+1}+\sqrt{2(n-1)}-1\leq 0 \\
\Leftrightarrow\quad & \left(\sqrt{\|\tilde{S}_{\rm Im}\|_F+1}-\sqrt{\frac{n-1}{2}}\right)^2-\frac{n-1}{2}+\sqrt{2(n-1)}-1\leq 0 \\
\Leftrightarrow\quad & \left(\sqrt{\|\tilde{S}_{\rm Im}\|_F+1}-\sqrt{\frac{n-1}{2}}\right)^2\leq\left(\sqrt{\frac{n-1}{2}}-1\right)^2 \\
\Leftrightarrow\quad & \sqrt{\frac{n-1}{2}}-\left(\sqrt{\frac{n-1}{2}}-1\right)\leq\sqrt{\|\tilde{S}_{\rm Im}\|_F+1}\leq\sqrt{\frac{n-1}{2}}+\left(\sqrt{\frac{n-1}{2}}-1\right) \\
\Leftrightarrow\quad & 1\leq\|\tilde{S}_{\rm Im}\|_F+1\leq(\sqrt{2(n-1)}-1)^2 \\
\Leftrightarrow\quad & 0\leq\|\tilde{S}_{\rm Im}\|_F\leq 2(n-1)-2\sqrt{2(n-1)},
\end{aligned} \tag{S99}$$

this always holds when $n\geq 5$ since from $S_{\rm Re}=I$ and $S_{\rm Re}+i\tilde{S}_{\rm Im}\geq 0$ we have $\|\tilde{S}_{\rm Im}\|_F\leq\|S_{\rm Re}\|_F=\sqrt{n}$, and $\sqrt{n}\leq 2(n-1)-2\sqrt{2(n-1)}$ when $n\geq 5$. Thus for $n\geq 5$ observables on pure states, the bound in Eq.(S92) is always tighter than the simple summation of Ozawa's relation for all pairs of the observables. We note that for pure states, $\mathcal{E}_A=\mathcal{E}_{Ozawa}$, the bound in Eq.(S92) is thus also tighter than the simple summation of $\mathcal{E}_A$ for all pairs.

S8. COMPARISON OF THE SDP BOUND FOR MORE THAN TWO OBSERVABLES WITH THE OZAWA'S BOUND ON MIXED STATES

For general n observables, $\{X_1,\cdots,X_n\}$, we can use the results on two observables to show that the SDP bound for $\sum_{i=1}^n\epsilon_i^2$ is always tighter than the bound obtained by simply summing up the Ozawa's relation for all $\epsilon_j^2+\epsilon_k^2$.

First note that for any pair of observables, we have $\mathcal{E}_0(X_j,X_k)\geq\sum_q\lambda_q\mathcal{E}_{|\phi_q\rangle}(X_j,X_k)$, and $\sum_q\lambda_q\mathcal{E}_{|\phi_q\rangle}(X_j,X_k)$ is tighter than the bound obtained from the Ozawa's relation. Thus the SDP bound for two observables is already tighter than the bound obtained from the Ozawa's relation. As $\sum_{i=1}^n\epsilon_i^2=\frac{1}{n-1}\sum_{j\neq k}(\epsilon_j^2+\epsilon_k^2)$, we can use the bound on two observables to obtain a bound on $\sum_{i=1}^n\epsilon_i^2$ as $\sum_{i=1}^n\epsilon_i^2\geq\frac{1}{n-1}\sum_{j\neq k}\mathcal{E}_0(X_j,X_k)$. This is tighter than the bound by summing up the Ozawa's relation for all pairs since each $\mathcal{E}_0(X_j,X_k)$ is tighter. To show the SDP bound for n observables is tighter than the bound obtained by summing up the Ozawa's relation, we just need to show

$$\mathcal{E}_0(X_1,\cdots,X_n)\geq\frac{1}{n-1}\sum_{j\neq k}\mathcal{E}_0(X_j,X_k). \tag{S100}$$

By taking $W=I_n$ with I_n as $n\times n$ Identity matrix, we have

$$\begin{aligned}
&\mathcal{E}_0(X_1,\cdots,X_n) \\
=&\min_{\mathbb{S},\{R_j\}_{j=1}^n}\ {\rm Tr}[(I\otimes\rho)(\mathbb{S}-\mathbb{R}\mathbb{X}^\dagger-\mathbb{X}\mathbb{R}^\dagger+\mathbb{X}\mathbb{X}^\dagger) \\
&\qquad |\mathbb{S}\geq\mathbb{R}\mathbb{R}^\dagger,\mathbb{S}_{jk}=\mathbb{S}_{kj}=\mathbb{S}_{jk}^\dagger,R_j=R_j^\dagger] \\
=&{\rm Tr}[(I\otimes\rho)(\mathbb{S}^*-\mathbb{R}^*\mathbb{X}^\dagger-\mathbb{X}\mathbb{R}^{*\dagger}+\mathbb{X}\mathbb{X}^\dagger)] \\
=&\frac{1}{n-1}\sum_{j,k}{\rm Tr}[(I^{(j,k)}\otimes\rho)(\mathbb{S}^*-\mathbb{R}^*\mathbb{X}^\dagger-\mathbb{X}\mathbb{R}^{*\dagger}+\mathbb{X}\mathbb{X}^\dagger)],
\end{aligned} \tag{S101}$$

here $\mathbb{S}^*$ and $\mathbb{R}^*$ denote the optimal operators that achieve the minimum for n observables, $I^{(j,k)}$ denote the diagonal matrix with its jth and kth diagonal element equal to 1 while others equal to 0. When $W = I^{(j,k)}$, only 2×2 blocks with indexes (j,k) contribute and we have

$$\begin{aligned}
&\mathrm{Tr}[(I^{(j,k)}\otimes\rho)(\mathbb{S}^*-\mathbb{R}^*\mathbb{X}^\dagger-\mathbb{X}\mathbb{R}^{*\dagger}+\mathbb{X}\mathbb{X}^\dagger)]\\
=&\mathrm{Tr}[(I_2\otimes\rho)(\mathbb{S}^{*(j,k)}-\mathbb{R}^{*(j,k)}\mathbb{X}^{\dagger(j,k)}\\
&\quad-\mathbb{X}^{(j,k)}\mathbb{R}^{*(j,k)\dagger}+\mathbb{X}^{(j,k)}\mathbb{X}^{\dagger(j,k)})]
\end{aligned}\tag{S102}$$

here I_2 is 2×2 Identity matrix, $\mathbb{S}^{*(j,k)} = \begin{pmatrix}\mathbb{S}^*_{jj} & \mathbb{S}^*_{jk}\\ \mathbb{S}^*_{kj} & \mathbb{S}^*_{kk}\end{pmatrix}$, $\mathbb{R}^{*(j,k)} = \begin{pmatrix}R^*_j & R^*_k\end{pmatrix}^\dagger$, $\mathbb{X}^{(j,k)} = \begin{pmatrix}X_j & X_k\end{pmatrix}^\dagger$. Note that $\mathbb{S}^*_{jk} = \mathbb{S}^*_{kj} = \mathbb{S}^{*\dagger}_{jk}$ for each (j,k), $R^*_j = R^{*\dagger}_j$ for each j, and from $\mathbb{S}^*\geq\mathbb{R}^*\mathbb{R}^{*\dagger}$, we have

$$\begin{pmatrix}\mathbb{S}^*_{jj} & \mathbb{S}^*_{jk}\\ \mathbb{S}^*_{kj} & \mathbb{S}^*_{kk}\end{pmatrix}\geq\begin{pmatrix}R^*_jR^*_j & R^*_jR^*_k\\ R^*_kR^*_j & R^*_kR^*_k\end{pmatrix},\ \text{for each } (j,k).\tag{S103}$$

Thus $\mathrm{Tr}[(I_2\otimes\rho)(\mathbb{S}^{*(j,k)}-\mathbb{R}^{*(j,k)}\mathbb{X}^{\dagger(j,k)}-\mathbb{X}^{(j,k)}\mathbb{R}^{*(j,k)\dagger}+\mathbb{X}^{(j,k)}\mathbb{X}^{\dagger(j,k)})]\geq\mathcal{E}_0(X_j,X_k)$ since $\mathcal{E}_0(X_j,X_k)=\min_{\mathbb{S},\mathbb{R}}\mathrm{Tr}[(I_2\otimes\rho)(\mathbb{S}-\mathbb{R}\mathbb{X}^\dagger-\mathbb{X}\mathbb{R}^\dagger+\mathbb{X}\mathbb{X}^\dagger)$, where $\mathbb{X}=\begin{pmatrix}X_j & X_k\end{pmatrix}^\dagger$ and the minimum is taken over all $\mathbb{S}=\begin{pmatrix}\mathbb{S}_{jj} & \mathbb{S}_{jk}\\ \mathbb{S}_{kj} & \mathbb{S}_{kk}\end{pmatrix}$ and $\mathbb{R}=\begin{pmatrix}R_j & R_k\end{pmatrix}^\dagger$ with $\mathbb{S}\geq\mathbb{R}\mathbb{R}^\dagger, \mathbb{S}_{jk}=\mathbb{S}_{kj}=\mathbb{S}^\dagger_{jk}, R_j=R^\dagger_j$. Thus

$$\begin{aligned}
&\mathcal{E}_0(X_1,\cdots,X_n)\\
=&\frac{1}{n-1}\sum_{j\neq k}\mathrm{Tr}[(I^{(j,k)}\otimes\rho)(\mathbb{S}^*-\mathbb{R}^*\mathbb{X}^\dagger-\mathbb{X}\mathbb{R}^{*\dagger}+\mathbb{X}\mathbb{X}^\dagger)]\\
\geq&\frac{1}{n-1}\sum_{j\neq k}\mathcal{E}_0(X_j,X_k).
\end{aligned}\tag{S104}$$

Since $\mathcal{E}_0(X_j,X_k)$ is tighter than the Ozawa's relation, the SDP bound is thus always tighter than the bound obtained from the simple summation of the Ozawa's relation for all pairs.

S9. EXAMPLES

In this section, we demonstrate the applications of (i) the bounds of $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$; (ii) $\mathcal{E}_A$ that can be analytically obtained and (iii) $\mathcal{E}_0$ that can be computed with SDP in specific examples.

A. Example: Error tradeoff relation for three observables on a qubit

Bounds obtained from $\mathrm{Tr}(S_{Re}^{-1}Q_{Re})$*:*

Consider the approximate measurement of the three Pauli operators $\frac{1}{2}\sigma_x,\ \frac{1}{2}\sigma_y,\ \frac{1}{2}\sigma_z$ for a general spin-1/2 state $\rho=\frac{1}{2}(I+r_x\sigma_x+r_y\sigma_y+r_z\sigma_z)$, where $|r_x|^2+|r_y|^2+|r_z|^2\leq 1$. In this case $X_1=\frac{1}{2}\sigma_x$, $X_2=\frac{1}{2}\sigma_y$, $X_3=\frac{1}{2}\sigma_z$, we have $(S_{\mathrm{Re}})_{jk}=\frac{1}{2}\mathrm{Tr}(\rho\{X_j,X_k\})=\frac{1}{2}\mathrm{Tr}(\rho\cdot\frac{1}{2}\delta_{jk}I)=\frac{1}{4}\delta_{jk}$ for $1\leq j,k\leq 3$. Thus $S_{\mathrm{Re}}=\frac{1}{4}I$, the error tradeoff relation is then given by

$$\sum_{j=1}^{3}\epsilon_j^2=\mathrm{Tr}(Q_{\mathrm{Re}})\geq\frac{1}{4}\left(\sqrt{4\|\tilde{S}_{\mathrm{Im}}\|_F+1}-1\right)^2,\tag{S105}$$

here $\tilde{S}_{\mathrm{Im}}$ is the imaginary part of $\tilde{S}=\sum_q\tilde{S}_{u_q}$ with each $\tilde{S}_{u_q}$ equals to either S_{u_q} or $\bar{S}_{u_q}$, where $(S_{u_q})_{jk}=\langle u_q|(\sqrt{\rho}\otimes\sqrt{\sigma})(X_jX_k\otimes I)(\sqrt{\rho}\otimes\sqrt{\sigma})|u_q\rangle$ with $\sigma=|\xi_0\rangle\langle\xi_0|$. The tightest bound is obtained by maximizing $\|\tilde{S}_{\mathrm{Im}}\|_F$ over all choices of $\{|u_q\rangle\}$ such that $\sum_q|u_q\rangle\langle u_q|=I$. Here instead of maximizing over all $\{|u_q\rangle\}$, we consider the maximization over $\{|u_1\rangle\otimes|\xi_l\rangle,|u_2\rangle\otimes|\xi_l\rangle|l=0,1,\cdots,d_A-1\}$, with $\{|u_1\rangle,|u_2\rangle\}$ and $\{|\xi_l\rangle|l=0,1,\cdots,d_A-1\}$ being orthonormal basis of the system and ancilla, respectively. With such choices we have

$$\begin{aligned}
(S_{u_{ql}})_{jk}&=\langle\xi_l|\langle u_q|(\sqrt{\rho}\otimes\sqrt{\sigma})(X_jX_k\otimes I)(\sqrt{\rho}\otimes\sqrt{\sigma})|u_q\rangle|\xi_l\rangle\\
&=\langle u_q|\sqrt{\rho}X_jX_k\sqrt{\rho}|u_q\rangle\delta_{0l}.
\end{aligned}\tag{S106}$$

The optimization is now over $\{|u_1\rangle, |u_2\rangle\}$ such that $|u_1\rangle\langle u_1| + |u_2\rangle\langle u_2| = I_S$ with I_S as the Identity on the system, together with the choices of taking the transpose on $(S_{u_q})_{jk} = \langle u_q|\sqrt{\rho}X_jX_k\sqrt{\rho}|u_q\rangle$, $q = 1, 2$. The optimization leads to a lower bound on the minimal error of approximating $\{X_j\}$ with any measurement on ρ. The bound could be further tightened by optimizing over all choices of $\{|u_q\rangle\}$ in the space of system and ancilla.

Before making the optimization, we emphasize that any choice leads to a valid bound. A simple choice is just to choose $\{|u_1\rangle, |u_2\rangle\}$ as $\{|0\rangle, |1\rangle\}$, and we are free to make the choices of taking the transpose on $\{S_{u_0}, S_{u_1}\}$. A direct calculation on $\{S_{u_0}, S_{u_1}\}$ gives

$$\max \|\tilde{S}_{\mathrm{Im}}\|_F^2 = \max\left\{\|S_{u_0,\mathrm{Im}} + S_{u_1,\mathrm{Im}}\|_F^2, \|S_{u_0,\mathrm{Im}} + S_{u_1,\mathrm{Im}}^T\|_F^2\right\} = \max\left\{\frac{1}{8}(r_x^2 + r_y^2 + r_z^2), \frac{1}{8}(1 - r_x^2 - r_y^2)\right\}, \tag{S107}$$

which leads to a tradeoff relation

$$\sum_{j=1}^{3} \epsilon_j^2 \geq \frac{1}{4}\left(\sqrt{4\max\|\tilde{S}_{\mathrm{Im}}\|_F + 1} - 1\right)^2 = \frac{1}{4}\left(\sqrt{4 \times \frac{1}{2\sqrt{2}}\max\left\{\sqrt{r_x^2 + r_y^2 + r_z^2}, \sqrt{1 - r_x^2 - r_y^2}\right\} + 1} - 1\right)^2. \tag{S108}$$

We now consider the maximization of $\|\tilde{S}_{\mathrm{Im}}\|_F$ over the choices of $\{|u_1\rangle, |u_2\rangle\}$. In this case we have

$$\max_{\{|u_q\rangle\}} \|\tilde{S}_{\mathrm{Im}}\|_F^2 = \max\left\{\sum_{jk}|w_{jk}|^2, \lambda_{\max}(S^{xx})\right\}, \tag{S109}$$

where $w_{jk} = \mathrm{Tr}(\mathcal{X}_{jk})$ with $\mathcal{X}_{jk} = \frac{1}{2i}\sqrt{\rho}[X_j, X_k]\sqrt{\rho}$, and $S^{xx} = \sum_{jk} x_{jk}x_{jk}^T$ with x_{jk} as a 3-dimensional vector with its rth entry given by $x_{jk}^{(r)} = \mathrm{Tr}(\mathcal{X}_{jk}\sigma_r)$, $1 \leq r \leq 3$, $\lambda_{\max}(S^{xx})$ denotes the largest eigenvalue of S^{xx}. We will show that for any $\rho = \frac{1}{2}(I + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z)$, we have $\sum_{jk}|w_{jk}|^2 = \frac{1}{8}\left(r_x^2 + r_y^2 + r_z^2\right) \leq \frac{1}{8}$ and $\lambda_{\max}(S^{xx}) = \frac{1}{8}$. This then leads to a bound as

$$\sum_{j=1}^{3} \epsilon_j^2 \geq \frac{1}{4}\left(\sqrt{4\max_{\{|u_q\rangle\}}\|\tilde{S}_{\mathrm{Im}}\|_F + 1} - 1\right)^2 = \frac{1}{4}\left(\sqrt{4 \times \frac{1}{2\sqrt{2}} + 1} - 1\right)^2 = \frac{1}{4}\left(\sqrt{\sqrt{2} + 1} - 1\right)^2, \tag{S110}$$

which is generally tighter than Eq.(S108).

We now sketch the procedure of the optimization to get the above bound. First note that the entries of $\tilde{S}_{\mathrm{Im}}$ are given by

$$(\tilde{S}_{\mathrm{Im}})_{jk} = \sum_{q=1}^{2}(\tilde{S}_{u_q,\mathrm{Im}})_{jk} = \sum_{q=1}^{2}(-1)^{s_q}\frac{1}{2i}\langle u_q|\sqrt{\rho}[X_j, X_k]\sqrt{\rho}|u_q\rangle, \tag{S111}$$

where $s_q = 0$ if $\tilde{S}_{u_q} = S_{u_q}$, $s_q = 1$ if $\tilde{S}_{u_q} = \bar{S}_{u_q}$. The Frobenius norm of $\tilde{S}_{\mathrm{Im}}$ can then be written as

$$\begin{aligned}
\|\tilde{S}_{\mathrm{Im}}\|_F^2 = \sum_{jk}|(\tilde{S}_{\mathrm{Im}})_{jk}|^2 &= \sum_{jk}\left|\sum_q(-1)^{s_q}\frac{1}{2i}\langle u_q|\sqrt{\rho}[X_j, X_k]\sqrt{\rho}|u_q\rangle\right|^2 \\
&= \sum_{jk}\left|\mathrm{Tr}\left(\left(\sum_q(-1)^{s_q}M_q\right)\frac{1}{2i}\sqrt{\rho}[X_j, X_k]\sqrt{\rho}\right)\right|^2 \\
&= \sum_{jk}\left|\mathrm{Tr}\left(\left(\sum_q(-1)^{s_q}M_q\right)\mathcal{X}_{jk}\right)\right|^2,
\end{aligned} \tag{S112}$$

here $M_q = |u_q\rangle\langle u_q|$ and $\mathcal{X}_{jk} = \frac{1}{2i}\sqrt{\rho}[X_j, X_k]\sqrt{\rho}$. The optimization can thus be reformulated as

$$\begin{aligned}
&\max_{\{M_q\}} \sum_{jk}\left|\mathrm{Tr}\left(\mathcal{X}_{jk}\sum_{q=1}^{2}(-1)^{s_q}M_q\right)\right|^2 \\
&\text{subject to } \sum_q M_q = I,\ M_q \geq 0,\ M_q^2 = M_q,
\end{aligned} \tag{S113}$$

where $s_q \in \{0,1\}, \forall q \in \{1,2\}$. For the qubit system, we introduce the Hermitian basis $\{\lambda_r\}_{r=0}^3 = \{\frac{1}{\sqrt{2}}I, \frac{1}{\sqrt{2}}\sigma_1, \frac{1}{\sqrt{2}}\sigma_2, \frac{1}{\sqrt{2}}\sigma_3\}$ such that $\mathrm{Tr}(\lambda_r\lambda_s) = \delta_{rs}$ for $0 \le r, s \le 3$. The Hermitian operators $\mathcal{X}_{jk}$ and M_q can then be vectorized as

$$\mathcal{X}_{jk} = \sum_{r=0}^{3} \mathrm{Tr}(\mathcal{X}_{jk}\lambda_r)\lambda_r = \frac{1}{2}\left[\mathrm{Tr}(\mathcal{X}_{jk})I + \sum_{r=1}^{3}\mathrm{Tr}(\mathcal{X}_{jk}\sigma_r)\sigma_r\right] = \frac{1}{2}\left(w_{jk}I + \sum_{r=1}^{3} x_{jk}^{(r)}\sigma_r\right), \tag{S114}$$

$$M_q = \sum_{r=0}^{3} \mathrm{Tr}(M_q\lambda_r)\lambda_r = \frac{1}{2}\left[\mathrm{Tr}(M_q)I + \sum_{r=1}^{3}\mathrm{Tr}(M_q\sigma_r)\sigma_r\right] = \frac{1}{2}\left(w_q I + \sum_{r=1}^{3} x_q^{(r)}\sigma_r\right), \tag{S115}$$

where we have defined $w_{jk} = \mathrm{Tr}(\mathcal{X}_{jk})$, $x_{jk}^{(r)} = \mathrm{Tr}(\mathcal{X}_{jk}\sigma_r)$, $w_q = \mathrm{Tr}(M_q) = 1$, $x_q^{(r)} = \mathrm{Tr}(M_q\sigma_r)$. Let $x_q = (x_q^{(1)}, x_q^{(2)}, x_q^{(3)})$ as a three-dimensional vectors, from $M_q \ge 0$, we can obtain $\sum_{r=1}^3 \left(x_q^{(r)}\right)^2 \le 1$, i.e., $\|x_q\| \le 1$, and from $M_q^2 = M_q$, we have $\|x_q\| = 1$. The optimization can thus be reformulated as

$$\begin{aligned} &\max_{\{x_q\}} \frac{1}{4}\sum_{jk}\left|w_{jk}\left[\sum_q (-1)^{s_q}\right] + \sum_{r=1}^{3} x_{jk}^{(r)}\left[\sum_q (-1)^{s_q} x_q^{(r)}\right]\right|^2 \\ &\text{subject to } \sum_q x_q = \mathbf{0},\ \|x_q\| = 1, \end{aligned} \tag{S116}$$

where $s_q \in \{0,1\}, \forall q \in \{1,2\}$.

When $s_1 = s_2 = 0$ or $s_1 = s_2 = 1$,

$$\max_{\{|u_q\rangle\}} \|\tilde{S}_{\mathrm{Im}}\|_F^2 = \max_{\{x_q\}} \frac{1}{4}\sum_{jk}\left|2w_{jk} + \sum_{r=1}^{3} x_{jk}^{(r)}\left(x_1^{(r)} + x_2^{(r)}\right)\right|^2 = \sum_{jk}|w_{jk}|^2. \tag{S117}$$

When $s_1 = 0$, $s_2 = 1$ or $s_1 = 1$, $s_2 = 0$,

$$\begin{aligned} \max_{\{|u_q\rangle\}} \|\tilde{S}_{\mathrm{Im}}\|_F^2 &= \max_{\{x_q\}} \frac{1}{4}\sum_{jk}\left|\sum_{r=1}^{3} x_{jk}^{(r)}\left(x_1^{(r)} - x_2^{(r)}\right)\right|^2 \\ &= \max_{\{x_1\}} \frac{1}{4}\sum_{jk}\left|2\sum_{r=1}^{3} x_{jk}^{(r)}x_1^{(r)}\right|^2 = \max_{\{x_1\}} \sum_{jk}\sum_{rs} x_{jk}^{(r)}x_1^{(r)}x_{jk}^{(s)}x_1^{(s)} \\ &= \max_{\{x_1\}} \sum_{rs} x_1^{(r)}\left(\sum_{jk} x_{jk}^{(r)}x_{jk}^{(s)}\right)x_1^{(s)} \\ &= \max_{\{x_1\}} x_1^T S^{xx} x_1 \\ &= \lambda_{\max}(S^{xx}). \end{aligned} \tag{S118}$$

where we have defined a 3×3 symmetric matrix S^{xx} with its rs-th entry as $(S^{xx})_{rs} = \sum_{jk} x_{jk}^{(r)}x_{jk}^{(s)}$, i.e., $S^{xx} = \sum_{jk} x_{jk}x_{jk}^T$. $\lambda_{\max}(S^{xx})$ then denotes the largest eigenvalue of the matrix S^{xx}. The optimal x_1 is given by the eigenvector of S^{xx} which corresponds to the maximal eigenvalue, which further gives the optimal choice of $\{|u_q\rangle\}$.

Combining the two cases together, we have

$$\max_{\{|u_q\rangle\}} \|\tilde{S}_{\mathrm{Im}}\|_F^2 = \max\left\{\sum_{jk}|w_{jk}|^2, \lambda_{\max}(S^{xx})\right\}. \tag{S119}$$

By substituting ρ and $X_1 = \frac{1}{2}\sigma_x$, $X_2 = \frac{1}{2}\sigma_y$, $X_3 = \frac{1}{2}\sigma_z$ in the definition of w_{jk}, we have

$$\sum_{jk}|w_{jk}|^2 = \sum_{jk}\left|\frac{1}{2i}\mathrm{Tr}(\rho[X_j, X_k])\right|^2 = \frac{1}{8}\left(r_x^2 + r_y^2 + r_z^2\right). \tag{S120}$$

To calculate $\lambda_{\max}(S^{xx})$, note that for $\rho=\frac{1}{2}(I+r_x\sigma_x+r_y\sigma_y+r_z\sigma_z)$, we can equivalently write $r_x=\lambda\sin\theta\cos\varphi$, $r_y=\lambda\sin\theta\sin\varphi$, and $r_z=\lambda\cos\theta$ with $|\lambda|\le 1$. ρ then has an eigendecomposition as $\rho=\frac{1+\lambda}{2}|\psi_1\rangle\langle\psi_1|+\frac{1-\lambda}{2}|\psi_2\rangle\langle\psi_2|$ with $|\psi_1\rangle=\begin{pmatrix} e^{-i\varphi}\cos\frac{\theta}{2} \\ \sin\frac{\theta}{2}\end{pmatrix}$, $|\psi_2\rangle=\begin{pmatrix} e^{-i\varphi}\sin\frac{\theta}{2} \\ -\cos\frac{\theta}{2}\end{pmatrix}$ and $\langle\psi_1|\psi_2\rangle=0$. Thus

$$\sqrt{\rho}=\sqrt{\frac{1+\lambda}{2}}|\psi_1\rangle\langle\psi_1|+\sqrt{\frac{1-\lambda}{2}}|\psi_2\rangle\langle\psi_2|=\frac{1}{2}\begin{pmatrix} p+q\cos\theta & qe^{-i\varphi}\sin\theta \\ qe^{i\varphi}\sin\theta & p-q\cos\theta\end{pmatrix}, \tag{S121}$$

where we have defined $p=\sqrt{\frac{1+\lambda}{2}}+\sqrt{\frac{1-\lambda}{2}}$, $q=\sqrt{\frac{1+\lambda}{2}}-\sqrt{\frac{1-\lambda}{2}}$. Then we have $x_{11}=x_{22}=x_{33}=\mathbf{0}$, and

$$\begin{aligned} x_{12}=-x_{21}&=\frac{1}{8}\begin{pmatrix} q^2\cos\varphi\sin 2\theta \\ q^2\sin\varphi\sin 2\theta \\ p^2+q^2\cos 2\theta\end{pmatrix}, \quad x_{13}=-x_{31}=-\frac{1}{8}\begin{pmatrix} q^2\sin 2\varphi\sin^2\theta \\ p^2-q^2(\cos^2\theta+\cos 2\varphi\sin^2\theta) \\ q^2\sin\varphi\sin 2\theta\end{pmatrix} \\ x_{23}=-x_{32}&=\frac{1}{8}\begin{pmatrix} p^2-q^2(\cos^2\theta-\cos 2\varphi\sin^2\theta) \\ q^2\sin 2\varphi\sin^2\theta \\ q^2\cos\varphi\sin 2\theta\end{pmatrix}. \end{aligned} \tag{S122}$$

A few lines of calculation then gives

$$S^{xx}=\sum_{j,k=1}^{3}x_{jk}x_{jk}^T=\frac{1}{8}\begin{pmatrix} 1-r_y^2-r_z^2 & r_xr_y & r_zr_x \\ r_xr_y & 1-r_z^2-r_x^2 & r_yr_z \\ r_zr_x & r_yr_z & 1-r_x^2-r_y^2\end{pmatrix}, \tag{S123}$$

where we have represented λ, θ, φ with the original notations r_x, r_y, r_z. The largest eigenvalue of S^{xx} is $\frac{1}{8}$, with corresponding eigenvectors $x_1=\frac{1}{\sqrt{r_x^2+r_y^2+r_z^2}}(r_x\ r_y\ r_z)^T$ for $r_x^2+r_y^2+r_z^2\neq 0$. In this case the optimal choice of $|u_1\rangle$ is the pure state with x_1 as its bloch vector, and the optimal $|u_2\rangle$ is the orthogonal complement of $|u_1\rangle$.

Thus for any $\rho=\frac{1}{2}(I+r_x\sigma_x+r_y\sigma_y+r_z\sigma_z)$, we have $\sum_{jk}|w_{jk}|^2=\frac{1}{8}\left(r_x^2+r_y^2+r_z^2\right)\le\frac{1}{8}$ and $\lambda_{\max}(S^{xx})=\frac{1}{8}$, Eq.(S105) and Eq.(S109) then lead to

$$\sum_{j=1}^{3}\epsilon_j^2\ge\frac{1}{4}\left(\sqrt{4\max_{\{|u_q\rangle\}}\|\tilde{S}_{\mathrm{Im}}\|_F+1}-1\right)^2=\frac{1}{4}\left(\sqrt{4\times\frac{1}{2\sqrt{2}}+1}-1\right)^2=\frac{1}{4}\left(\sqrt{\sqrt{2}+1}-1\right)^2. \tag{S124}$$

This bounds the minimal error of approximately implementing $\{\frac{1}{2}\sigma_x,\frac{1}{2}\sigma_y,\frac{1}{2}\sigma_z\}$ with a single measurement on ρ.

Bounds obtained from $\mathcal{E}_A$ and the SDP $\mathcal{E}_0$:

Next we calculate the analytical bound $\mathcal{E}_A$ for each pairs of the observables. For the ease of comparison, here we choose specific values of $\theta=\frac{\pi}{2}$, $\phi=0$, such that $\rho=\frac{1}{2}(I+\lambda\sigma_x)$. For the simultaneous measurement of (X_1,X_2), it is straightforward to obtain that

$$\sqrt{\rho}[X_1,X_2]\sqrt{\rho}=\frac{i}{4}\sqrt{1-\lambda^2}\begin{pmatrix} 1 & 0 \\ 0 & -1\end{pmatrix}. \tag{S125}$$

By choosing $\{|u_q\rangle\}$ as the eigenvectors of $\sqrt{\rho}[X_1,X_2]\sqrt{\rho}$, i.e., $\{|u_q\rangle\}=\{|0\rangle,|1\rangle\}$, the bound $\mathcal{E}_A$ can be directly calculated as

$$\epsilon_1^2+\epsilon_2^2\ge\mathcal{E}_A(X_1,X_2)=\frac{1}{4}(1-\lambda^2). \tag{S126}$$

Similarly for the simultaneous measurement of (X_2,X_3) and (X_3,X_1), the bound $\mathcal{E}_A$ can be calculated as

$$\begin{aligned} \epsilon_2^2+\epsilon_3^2&\ge\mathcal{E}_A(X_2,X_3)=\frac{1}{4}, \\ \epsilon_3^2+\epsilon_1^2&\ge\mathcal{E}_A(X_3,X_1)=\frac{1}{4}(1-\lambda^2). \end{aligned} \tag{S127}$$

For the simultaneous measurement of (X_1,X_2,X_3), a direct summation of the bounds for each pairs of the observables gives

$$\epsilon_1^2+\epsilon_2^2+\epsilon_3^2\ge\frac{1}{2}(\mathcal{E}_A(X_1,X_2)+\mathcal{E}_A(X_2,X_3)+\mathcal{E}_A(X_3,X_1))=\frac{1}{8}+\frac{1}{4}(1-\lambda^2). \tag{S128}$$

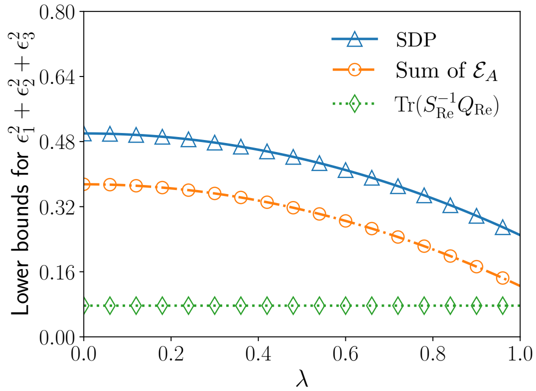

Figure S3. Lower bounds for the errors of the simultaneous measurement of (X_1, X_2, X_3) given by (i) bound that obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$; (ii) the simple summation of the analytical bound $\mathcal{E}_A$ for all pairs and (iii) $\mathcal{E}_0$ that can be computed with SDP.

The SDP $\mathcal{E}_0$ also gives a bound for (X_1, X_2, X_3) directly, which can be computed efficiently. In Fig. S3, we make comparisons among the three bounds, (i) bound that obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$; (ii) the simple summation of the analytical bound $\mathcal{E}_A$ for all pairs and (iii) $\mathcal{E}_0$ that can be computed with SDP. It can be seen that the SDP bound is the tightest among the three bounds. We note that although the simple summation of $\mathcal{E}_A$ for each pairs of the observables is tighter than $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$ in this example, there are cases the bound obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$ can be tighter, as we have shown in the previous section.

B. Example: Tradeoff relation for the estimation of multiple parameters in a qubit

Bounds obtained from $\mathrm{Tr}(S_{Re}^{-1}Q_{Re})$*:*

The error-tradeoff relation for multiple observables can be used to obtain tradeoff relation in multiparameter quantum estimation. Here we consider the simultaneous estimation of the three parameters $(\lambda, \theta, \varphi)$ in $\rho = \frac{1}{2}(I + r_x\sigma_x + r_y\sigma_y + r_z\sigma_z)$, where $r_x = \lambda \sin\theta\cos\varphi$, $r_y = \lambda\sin\theta\sin\varphi$, and $r_z = \lambda\cos\theta$, here $|\lambda| \le 1$.

For any measurement on ρ, the classical Fisher information matrix, F_C, is bounded as

$$\mathrm{Tr}[F_Q^{-1}F_C] \le 3 - \left(\sqrt{\|F_Q^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}}F_Q^{-\frac{1}{2}}\|_F + 1} - 1\right)^2, \tag{S129}$$

The SLDs for $(\lambda, \theta, \varphi)$ are given by

$$L_\lambda = \frac{1}{1-\lambda^2}(\vec{n}\cdot\vec{\sigma} - \lambda I), \quad L_\theta = \lambda(\partial_\theta\vec{n}\cdot\vec{\sigma}), \quad L_\varphi = \lambda(\partial_\varphi\vec{n}\cdot\vec{\sigma}), \tag{S130}$$

here $\vec{n} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$, $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, and the quantum Fisher information matrix is

$$F_Q = \begin{pmatrix} \frac{1}{1-\lambda^2} & 0 & 0 \\ 0 & \lambda^2 & 0 \\ 0 & 0 & \lambda^2\sin^2\theta \end{pmatrix}. \tag{S131}$$

We can use Eq.(S109) to maximize $\tilde{S}_{\mathrm{Im}}$ over $\{|u_q\rangle\}$. Note that under the reparameterization which makes $F_Q = I$, we have $F_Q^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}}F_Q^{-\frac{1}{2}} = \tilde{S}_{\mathrm{Im}}$. The SLD operators under this reparametrization are given by $\tilde{L}_k = \sum_j (F_Q^{-\frac{1}{2}})_{jk}L_j$, i.e.,

$$\tilde{L}_\lambda = \frac{1}{\sqrt{1-\lambda^2}}(\vec{n}\cdot\vec{\sigma} - \lambda I), \quad \tilde{L}_\theta = \partial_\theta\vec{n}\cdot\vec{\sigma}, \quad \tilde{L}_\varphi = \frac{1}{|\sin\theta|}(\partial_\varphi\vec{n}\cdot\vec{\sigma}). \tag{S132}$$

From Eq.(S121) we can obtain $\sqrt{\rho}=\frac{1}{2}(pI+q\vec{n}\cdot\sigma)$, thus we have

$$\begin{aligned}\mathcal{X}_{12}&=\frac{1}{2i}\sqrt{\rho}[\tilde{L}_\lambda,\tilde{L}_\theta]\sqrt{\rho}=\frac{1}{2\sin\theta}(\partial_\varphi\vec{n}\cdot\vec{\sigma}),\\ \mathcal{X}_{23}&=\frac{1}{2i}\sqrt{\rho}[\tilde{L}_\theta,\tilde{L}_\varphi]\sqrt{\rho}=\frac{|\sin\theta|}{2\sin\theta}\left(\lambda I+\vec{n}\cdot\vec{\sigma}\right),\\ \mathcal{X}_{31}&=\frac{1}{2i}\sqrt{\rho}[\tilde{L}_\varphi,\tilde{L}_\lambda]\sqrt{\rho}=\frac{|\sin\theta|}{2\sin\theta}\left(\partial_\theta\vec{n}\cdot\vec{\sigma}\right),\end{aligned}\tag{S133}$$

which gives

$$x_{12}=-x_{21}=\frac{1}{\sin\theta}\partial_\varphi\vec{n},\quad x_{23}=-x_{32}=\frac{|\sin\theta|}{\sin\theta}\vec{n},\quad x_{13}=-x_{31}=-\frac{|\sin\theta|}{\sin\theta}\partial_\theta\vec{n}.\tag{S134}$$

Note that $x_{11}=x_{22}=x_{33}=\mathbf{0}$, it is then easy to obtain $S^{xx}=\sum_{j,k=1}^{3}x_{jk}x_{jk}^T=2I$. In this case $\{|u_q\rangle\}$ can be chosen arbitrarily and we always have $\lambda_{\max}(S^{xx})=2$. It is also straightforward to obtain $\sum_{jk}|w_{jk}|^2=\sum_{jk}|\mathrm{Tr}(\mathcal{X}_{jk})|^2=2\lambda^2\leq 2$, thus

$$\max_{\{|u_q\rangle\}}\|F_Q^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}}F_Q^{-\frac{1}{2}}\|_F=\sqrt{2}.\tag{S135}$$

This leads to a tradeoff relation for the estimation of $\{\lambda,\theta,\varphi\}$ as

$$\mathrm{Tr}[F_Q^{-1}F_C]\leq 3-\left(\sqrt{\max_{\{|u_q\rangle\}}\|F_Q^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}}F_Q^{-\frac{1}{2}}\|_F+1}-1\right)^2=3-\left(\sqrt{\sqrt{2}+1}-1\right)^2.\tag{S136}$$

Bounds obtained from $\mathcal{E}_A$ and the SDP $\mathcal{E}_0$:

By doing summation over the analytical bound $\mathcal{E}_A$ for each pairs of $(\tilde{L}_j,\tilde{L}_k)$, we directly have

$$\mathrm{Tr}[F_Q^{-1}F_C]\leq n-\frac{1}{2(n-1)}\|C\|_F^2,\tag{S137}$$

with the jk-th entry of C given as $(C)_{jk}=\sqrt{\mathcal{E}_A(\tilde{L}_j,\tilde{L}_k)}$. Again for the ease of comparison, we choose specific values of $\theta=\frac{\pi}{2}$, $\phi=0$. Substituting $\tilde{L}_\lambda$, $\tilde{L}_\theta$, $\tilde{L}_\varphi$ in $\mathcal{E}_A(\tilde{L}_j,\tilde{L}_k)$, we directly have $\mathcal{E}_A(\tilde{L}_\lambda,\tilde{L}_\theta)=\mathcal{E}_A(\tilde{L}_\theta,\tilde{L}_\varphi)=\mathcal{E}_A(\tilde{L}_\varphi,\tilde{L}_\lambda)=1$, and

$$\mathrm{Tr}[F_Q^{-1}F_C]\leq 3-\frac{1}{4}\|C\|_F^2=\frac{3}{2}.\tag{S138}$$

By substituting $\tilde{L}_\lambda$, $\tilde{L}_\theta$, $\tilde{L}_\varphi$, the SDP $\mathcal{E}_0$ also gives a bound directly as $\mathrm{Tr}[F_Q^{-1}F_C]\leq n-\mathcal{E}_0=1$, which is the tightest among all bounds and coincides with the Gill-Massar bound for qubit states.

C. Example: Tradeoff relation under collective measurement

While the tradeoff relation is obtained under local measurements, it can also be used to obtain tradeoff relations for collective measurements on p copies of the states by simply replacing ρ with $\rho_x^{\otimes p}$.

As an example we consider the simultaneous estimation of the three parameters (λ,θ,φ) in a state $\rho=\frac{1}{2}(I+r_x\sigma_x+r_y\sigma_y+r_z\sigma_z)$, where $r_x=\lambda\sin\theta\cos\varphi$, $r_y=\lambda\sin\theta\sin\varphi$, and $r_z=\lambda\cos\theta$, $|\lambda|\leq 1$. But different from previous section, here we also allow collective measurement on two-copies of the state, $\rho\otimes\rho$.

To obtain the tradeoff relation under the collective measurement on 2-copies of the state, we can simply treat $\rho\otimes\rho$ as a larger single state. Its SLDs are given by $L_{j2}=L_j\otimes I+I\otimes L_j$ with $j=\lambda,\theta,\varphi$ and the quantum Fisher information matrix is $F_{Q2}=2F_Q$ with F_Q as the quantum Fisher information matrix of a single ρ.

Bounds obtained from $\mathrm{Tr}(S_{Re}^{-1}Q_{Re})$:

If we choose $\{|u_q\rangle\}$ as the computational basis with $|u_0\rangle = |00\rangle$, $|u_1\rangle = |01\rangle$, $|u_2\rangle = |10\rangle$, $|u_3\rangle = |11\rangle$, the corresponding $S_{u_q,\mathrm{Im}2}$ are given by

$$\begin{aligned}
S_{u_0,\mathrm{Im}2} &= \begin{pmatrix} 0 & 0 & -\frac{\lambda(1+\lambda\cos\theta)\sin^2\theta}{2\sqrt{1-\lambda^2}} \\ 0 & 0 & -\frac{1}{2}\lambda^2\sin\theta(1+\lambda\cos\theta)(\cos\theta+\lambda) \\ \frac{\lambda(1+\lambda\cos\theta)\sin^2\theta}{2\sqrt{1-\lambda^2}} & \frac{1}{2}\lambda^2\sin\theta(1+\lambda\cos\theta)(\cos\theta+\lambda) & 0 \end{pmatrix} \\
S_{u_1,\mathrm{Im}2} = S_{u_2,\mathrm{Im}2} &= \begin{pmatrix} 0 & 0 & \frac{\lambda^2\cos\theta\sin^2\theta}{2\sqrt{1-\lambda^2}} \\ 0 & 0 & -\frac{1}{2}\lambda^3\sin^3\theta \\ -\frac{\lambda^2\cos\theta\sin^2\theta}{2\sqrt{1-\lambda^2}} & \frac{1}{2}\lambda^3\sin^3\theta & 0 \end{pmatrix} \\
S_{u_3,\mathrm{Im}2} &= \begin{pmatrix} 0 & 0 & \frac{\lambda(1-\lambda\cos\theta)\sin^2\theta}{2\sqrt{1-\lambda^2}} \\ 0 & 0 & \frac{1}{2}\lambda^2\sin\theta(1-\lambda\cos\theta)(\cos\theta-\lambda) \\ -\frac{\lambda(1-\lambda\cos\theta)\sin^2\theta}{2\sqrt{1-\lambda^2}} & -\frac{1}{2}\lambda^2\sin\theta(1-\lambda\cos\theta)(\cos\theta-\lambda) & 0 \end{pmatrix}
\end{aligned} \tag{S139}$$

The optimal $\|F_{Q2}^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}2}F_{Q2}^{-\frac{1}{2}}\|_F$ is thus given by the maximum of the following values,

$$\begin{aligned}
\|F_{Q2}^{-\frac{1}{2}}(S_{u_0,\mathrm{Im}2}+S_{u_1,\mathrm{Im}2}+S_{u_2,\mathrm{Im}2}+S_{u_3,\mathrm{Im}2})F_{Q2}^{-\frac{1}{2}}\|_F &= \sqrt{2}|\lambda|, \\
\|F_{Q2}^{-\frac{1}{2}}(S_{u_0,\mathrm{Im}2}+S_{u_1,\mathrm{Im}2}^T+S_{u_2,\mathrm{Im}2}^T+S_{u_3,\mathrm{Im}2})F_{Q2}^{-\frac{1}{2}}\|_F &= \sqrt{2}|\lambda\cos\theta|, \\
\|F_{Q2}^{-\frac{1}{2}}(S_{u_0,\mathrm{Im}2}+S_{u_1,\mathrm{Im}2}+S_{u_2,\mathrm{Im}2}+S_{u_3,\mathrm{Im}2}^T)F_{Q2}^{-\frac{1}{2}}\|_F &= \sqrt{\frac{1}{2}(1+\lambda^2)(1+\lambda^2\cos^2\theta)+\lambda^3\cos\theta\sin^2\theta}, \\
\|F_{Q2}^{-\frac{1}{2}}(S_{u_0,\mathrm{Im}2}+S_{u_1,\mathrm{Im}2}^T+S_{u_2,\mathrm{Im}2}^T+S_{u_3,\mathrm{Im}2}^T)F_{Q2}^{-\frac{1}{2}}\|_F &= \sqrt{\frac{1}{2}(1+\lambda^2)(1+\lambda^2\cos^2\theta)-\lambda^3\cos\theta\sin^2\theta}.
\end{aligned} \tag{S140}$$

By simply replacing F_Q and $\tilde{S}_{\mathrm{Im}}$ with F_{Q2} and $\tilde{S}_{\mathrm{Im}2}$ in Eq.(S129), we can get

$$\begin{aligned}
\frac{1}{2}\mathrm{Tr}[F_Q^{-1}F_{C2}] &\leq 3-\left(\sqrt{\|F_{Q2}^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}2}F_{Q2}^{-\frac{1}{2}}\|_F+1}-1\right)^2 \\
&= 3-\left(\sqrt{\max\left\{\sqrt{2}|\lambda|,\sqrt{\frac{1}{2}(1+\lambda^2)(1+\lambda^2\cos^2\theta)+|\lambda^3\cos\theta\sin^2\theta|}\right\}+1}-1\right)^2,
\end{aligned} \tag{S141}$$

here F_{C2} denotes the classical Fisher information matrix under the collective measurement on $\rho\otimes\rho$, $\frac{1}{2}F_{C2}$ can be regarded as the average Fisher information on each ρ. This bounds the achievable precision limit under the collective measurements on 2 copies of the quantum state. Specifically for $\theta=\frac{\pi}{2}$, we have

$$\frac{1}{2}\mathrm{Tr}[F_Q^{-1}F_{C2}] \leq 3-\left(\sqrt{\max\left\{\sqrt{2}|\lambda|,\sqrt{\frac{1}{2}(1+\lambda^2)}\right\}+1}-1\right)^2. \tag{S142}$$

Bounds obtained from $\mathcal{E}_A$ and the SDP $\mathcal{E}_0$:

Under the reparameterization that $F_{Q2}=2F_Q=I$, by replacing (X_j,X_k) with $(\tilde{L}_{j2},\tilde{L}_{k2})$, the bound $\mathcal{E}_A$ gives

$$\epsilon_j^2+\epsilon_k^2 = 1-(F_{C2})_{jj}+1-(F_{C2})_{kk} \geq \mathcal{E}_A(\tilde{L}_{j2},\tilde{L}_{k2}). \tag{S143}$$

By doing summations over each pairs of (j,k), we directly have

$$\frac{1}{2}\mathrm{Tr}[F_Q^{-1}F_{C2}] \leq n-\frac{1}{2(n-1)}\|C_2\|_F^2, \tag{S144}$$

with the jk-th entry of C_2 given as $(C_2)_{jk}=\sqrt{\mathcal{E}_A(\tilde{L}_{j2},\tilde{L}_{k2})}$. Again for the ease of comparison, we choose specific values of $\theta=\frac{\pi}{2}$, $\phi=0$. Substituting $\tilde{L}_{\lambda 2}$, $\tilde{L}_{\theta 2}$, $\tilde{L}_{\varphi 2}$ in $\mathcal{E}_A(\tilde{L}_{j2},\tilde{L}_{k2})$, we can directly obtain a bound. By substituting $\tilde{L}_{\lambda 2}$, $\tilde{L}_{\theta 2}$, $\tilde{L}_{\varphi 2}$, the SDP $\mathcal{E}_0$ also gives a bound directly as $\frac{1}{2}\mathrm{Tr}[F_Q^{-1}F_{C2}]\leq n-\mathcal{E}_0$. Fig. S4 then illustrate the comparison among the three bounds with λ varies from 0 to 1.

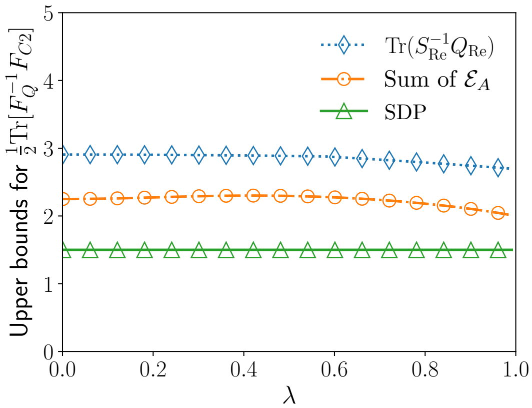

Figure S4. Upper bounds for $\frac{1}{2}\mathrm{Tr}[F_Q^{-1}F_{C2}]$ given by (i) bound that obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$; (ii) the simple summation of the analytical bound $\mathcal{E}_A$ for all pairs and (iii) $\mathcal{E}_0$ that can be computed with SDP.

D. Example: Tradeoff relation for the estimation of multiple parameters in three qubits

We provide an example with three qubits to show that the analytical bound obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$ can be tighter than the simple summation over the bounds $\mathcal{E}_A$ for each pairs of observables.

We consider a parameterized pure state $|\psi\rangle = \sin\theta_0|\psi_1\rangle + \cos\theta_0 e^{i\phi_0}|\psi_2\rangle$, where

$$\begin{aligned}
|\psi_1\rangle &= \sin\theta_1\sin\theta_2|001\rangle + \sin\theta_1\cos\theta_2 e^{i\phi_1}|010\rangle + \cos\theta_1 e^{i\phi_2}|100\rangle,\\
|\psi_2\rangle &= \sin\theta_3\sin\theta_4|110\rangle + \sin\theta_3\cos\theta_4 e^{i\phi_3}|101\rangle + \cos\theta_3 e^{i\phi_4}|011\rangle.
\end{aligned} \tag{S145}$$

We first consider the estimation of $\{\theta_0,\theta_1,\theta_2,\phi_1,\phi_2\}$ in $|\psi\rangle = \sin\theta_0|\psi_1\rangle + \cos\theta_0 e^{i\phi_0}|\psi_2\rangle$. In this case $F_Q = \begin{pmatrix} F_{Q1} & \mathbf{0} \\ \mathbf{0} & F_{Q2}\end{pmatrix}$ with

$$\begin{aligned}
F_{Q1} &= \mathrm{diag}\{4, 4\sin^2\theta_0, 4\sin^2\theta_0\sin^2\theta_1\},\\
F_{Q2} &= \begin{pmatrix} 4\sin^2\theta_0\sin^2\theta_1\cos^2\theta_2(1-\sin^2\theta_0\sin^2\theta_1\cos^2\theta_2) & -\sin^4\theta_0\sin^2 2\theta_1\cos^2\theta_2 \\ -\sin^4\theta_0\sin^2 2\theta_1\cos^2\theta_2 & 4\sin^2\theta_0\cos^2\theta_1(1-\sin^2\theta_0\cos^2\theta_1)\end{pmatrix},
\end{aligned} \tag{S146}$$

and $S_{\mathrm{Im}} = \begin{pmatrix} \mathbf{0} & S_{\mathrm{Im}1}^\dagger \\ S_{\mathrm{Im}1} & \mathbf{0}\end{pmatrix}$ where

$$S_{\mathrm{Im}1} = 2\begin{pmatrix} \sin 2\theta_0\sin^2\theta_1\cos^2\theta_2 & \sin^2\theta_0\sin 2\theta_1\cos^2\theta_2 & -\sin^2\theta_0\sin^2\theta_1\sin 2\theta_2 \\ \sin 2\theta_0\cos^2\theta_1 & -\sin^2\theta_0\sin 2\theta_1 & 0\end{pmatrix}. \tag{S147}$$

Thus

$$F_Q^{-\frac{1}{2}}S_{\mathrm{Im}}F_Q^{-\frac{1}{2}} = \begin{pmatrix} F_{Q1}^{-\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & F_{Q2}^{-\frac{1}{2}}\end{pmatrix}\begin{pmatrix} \mathbf{0} & S_{\mathrm{Im}1}^\dagger \\ S_{\mathrm{Im}1} & \mathbf{0}\end{pmatrix}\begin{pmatrix} F_{Q1}^{-\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & F_{Q2}^{-\frac{1}{2}}\end{pmatrix} = \begin{pmatrix} \mathbf{0} & F_{Q1}^{-\frac{1}{2}}S_{\mathrm{Im}1}^\dagger F_{Q2}^{-\frac{1}{2}} \\ F_{Q2}^{-\frac{1}{2}}S_{\mathrm{Im}1}F_{Q1}^{-\frac{1}{2}} & \mathbf{0}\end{pmatrix}, \tag{S148}$$

and

$$\|F_Q^{-\frac{1}{2}}S_{\mathrm{Im}}F_Q^{-\frac{1}{2}}\|_F^2 = 2\|F_{Q2}^{-\frac{1}{2}}S_{\mathrm{Im}1}F_{Q1}^{-\frac{1}{2}}\|_F^2 = 2\mathrm{Tr}\left(F_{Q2}^{-1}S_{\mathrm{Im}1}F_{Q1}^{-1}S_{\mathrm{Im}1}^\dagger\right) = 4. \tag{S149}$$

We then have

$$\mathrm{Tr}[F_Q^{-1}F_C] \leq n - \left(\sqrt{\|F_Q^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}}F_Q^{-\frac{1}{2}}\|_F + 1} - 1\right)^2 = 5 - (\sqrt{3}-1)^2 \approx 4.464. \tag{S150}$$

Compared with the simple summation of the analytical bound for pairs of the observables, which cannot be tighter than $n - \frac{1}{2(n-1)}\|F_Q^{-\frac{1}{2}}S_{\mathrm{Im}}F_Q^{-\frac{1}{2}}\|_F^2 = 5 - \frac{1}{8}\times 4 = \frac{9}{2}$, the bound obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$ is strictly tighter. The tight

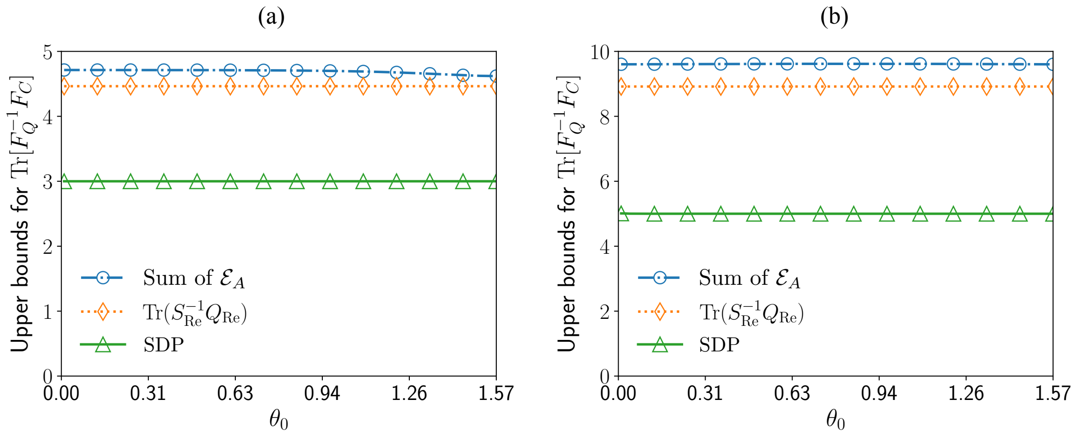

Figure S5. Upper bounds for $\mathrm{Tr}[F_Q^{-1}F_C]$ of (a) estimation of $n=5$ parameters $\{\theta_0,\theta_1,\theta_2,\phi_1,\phi_2\}$, (b) estimation of $n=10$ parameters $\{\theta_0,\theta_1,\theta_2,\theta_3,\theta_4,\phi_0,\phi_1,\phi_2,\phi_3,\phi_4\}$. Here the bounds are given by (i) bound that obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$; (ii) the simple summation of the analytical bound $\mathcal{E}_A$ for all pairs and (iii) $\mathcal{E}_0$ that can be computed with SDP.

bound can be obtained by computing the SDP $\mathcal{E}_0$. Specifically, for $\theta_j=\pi/4$, $1\le j\le 4$ and $\phi_k=\pi/4$, $0\le k\le 4$, we compare the three bounds with θ_0 varies from 0 to $\pi/2$. As illustrated in Fig. S5(a), the analytical bound obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$ is tighter than the simple summation of $\mathcal{E}_A$ for all pairs, while the tight bound is given by $\mathcal{E}_0$ that can be computed with SDP, which is approximately $\mathrm{Tr}[F_Q^{-1}F_C]\le 3$.

For the estimation of the full set, $\{\theta_0,\theta_1,\theta_2,\theta_3,\theta_4,\phi_0,\phi_1,\phi_2,\phi_3,\phi_4\}$, we can obtain the bound numerically. Specifically, for $\theta_j=\pi/4$ and $\phi_k=\pi/4$, $0\le j,k\le 4$, we have $F_Q=\begin{pmatrix}F_{Q1} & \mathbf{0}\\ \mathbf{0} & F_{Q2}\end{pmatrix}$ where

$$F_{Q1}=\mathrm{diag}\{4,2,1,2,1\},F_{Q2}=\begin{pmatrix}1 & -\frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & \frac{1}{2}\\ -\frac{1}{4} & \frac{7}{16} & -\frac{1}{8} & -\frac{1}{16} & -\frac{1}{8}\\ -\frac{1}{2} & -\frac{1}{8} & \frac{3}{4} & -\frac{1}{8} & -\frac{1}{4}\\ \frac{1}{4} & -\frac{1}{16} & -\frac{1}{8} & \frac{7}{16} & -\frac{1}{8}\\ \frac{1}{2} & -\frac{1}{8} & -\frac{1}{4} & -\frac{1}{8} & \frac{3}{4}\end{pmatrix}, \tag{S151}$$

and $S_{\mathrm{Im}}=\begin{pmatrix}\mathbf{0} & S_{\mathrm{Im1}}^{\dagger}\\ S_{\mathrm{Im1}} & \mathbf{0}\end{pmatrix}$ with

$$S_{\mathrm{Im1}}=\begin{pmatrix}-2 & 0 & 0 & 0 & 0\\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & 0 & 0\\ 1 & -1 & 0 & 0 & 0\\ -\frac{1}{2} & 0 & 0 & \frac{1}{2} & -\frac{1}{2}\\ -1 & 0 & 0 & -1 & 0\end{pmatrix}. \tag{S152}$$

We then have

$$\|F_Q^{-\frac{1}{2}}S_{\mathrm{Im}}F_Q^{-\frac{1}{2}}\|_F^2=2\|F_{Q2}^{-\frac{1}{2}}S_{\mathrm{Im1}}F_{Q1}^{-\frac{1}{2}}\|_F^2=2\mathrm{Tr}\left(F_{Q2}^{-1}S_{\mathrm{Im1}}F_{Q1}^{-1}S_{\mathrm{Im1}}^{\dagger}\right)=10, \tag{S153}$$

and

$$\mathrm{Tr}[F_Q^{-1}F_C]\le n-\left(\sqrt{\|F_Q^{-\frac{1}{2}}\tilde{S}_{\mathrm{Im}}F_Q^{-\frac{1}{2}}\|_F+1}-1\right)^2=10-\left(\sqrt{\sqrt{10}+1}-1\right)^2\approx 8.918. \tag{S154}$$

Compared with the simple summation of the analytical bound for pairs of the observables, which cannot be tighter than $n-\frac{1}{2(n-1)}\|F_Q^{-\frac{1}{2}}S_{\mathrm{Im}}F_Q^{-\frac{1}{2}}\|_F^2=10-\frac{1}{18}\times 10=\frac{85}{9}\approx 9.444$, the bound obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$ is strictly tighter. The tight bound can be obtained by computing the SDP $\mathcal{E}_0$. For $\theta_j=\pi/4$, $1\le j\le 4$ and $\phi_k=\pi/4$, $0\le k\le 4$, we compare the three bounds with θ_0 varies from 0 to $\pi/2$. As illustrated in Fig. S5(b), the analytical bound obtained from $\mathrm{Tr}(S_{\mathrm{Re}}^{-1}Q_{\mathrm{Re}})$ is tighter than the simple summation of $\mathcal{E}_A$ for all pairs, while the tight bound is given by $\mathcal{E}_0$ that can be computed with SDP, which is approximately $\mathrm{Tr}[F_Q^{-1}F_C]\le 5$.