Supplementary Information for "Optimal Provable Robustness of Quantum Classification via Quantum Hypothesis Testing"

Here, we provide detailed proofs for the robustness bounds presented in the paper in terms of the fidelity and trace distance, stated in Lemma 1, Corollary 1 and the bound for depolarized inputs (Eq. (43) in the main part). To that end, we first show a collection of technical lemmas related to quantum hypothesis testing with the goal of explicitly constructing Helstrom operators which attain a preassigned level of type-I error probability. These constructions of the Helstrom operators will then be used to derive an expression for the SDP (Eq. (3) in the main part) in terms of the fidelity between the input states.

Supplementary Note 1. TECHNICAL LEMMAS

Preliminaries. We first recall the central quantities of interest. As in the main part of this paper, the null hypothesis corresponds to a benign input state and is described by a density operator $\sigma \in \mathcal{S}(\mathcal{H})$ acting on a Hilbert space $\mathcal{H}$ of finite dimension $d := \dim(\mathcal{H}) < \infty$. The density operator for the alternative hypothesis is denoted by ρ and corresponds the adversarial state in the classification setting. A quantum hypothesis test is defined by a positive semi-definite operator $0 \leq M \leq \mathbb{1}_d$ and the type-I and type-II error probabilities associated with M are denoted by α and β and are defined by

$$\alpha(M;\, \sigma) := \mathrm{Tr}\left[\sigma M\right] \qquad \text{(type-I error)} \tag{1}$$

$$\beta(M;\, \rho) := \mathrm{Tr}\left[\rho(\mathbb{1} - M)\right] \qquad \text{(type-II error)} \tag{2}$$

Throughout this supplementary information we will omit the explicit dependence on σ and ρ whenever it is clear from context. For two Hermitian operators A and B, we write $A \geq B$ ($A \leq B$) if $A - B$ is positive (negative) semi-definite and $A > B$ ($A < B$) if $A - B$ is positive (negative) definite. For a Hermitian operator A with spectral decomposition $A = \sum_i \lambda_i P_i$ with eigenvalues $\{\lambda_i\}_i$ and orthogonal projections onto the associated eigenspaces $\{P_i\}_i$, we write

$$\{A > 0\} := \sum_{i:\, \lambda_i > 0} P_i, \;\; \{A < 0\} := \sum_{i:\, \lambda_i < 0} P_i \tag{3}$$

for the projections onto the eigenspaces associated with positive and negative eigenvalues respectively. Finally, for $t \geq 0$ define the operators

$$P_{t,+} := \{\rho - t\sigma > 0\}, \;\; P_{t,-} := \{\rho - t\sigma < 0\}, \;\; P_{t,0} := \mathbb{1} - P_{t,+} - P_{t,-}. \tag{4}$$

Helstrom operators are then defined as

$$M_t := P_{t,+} + X_t, \;\; 0 \leq X_t \leq P_{t,0}. \tag{5}$$

Finally, recall the SDP β^* (Eq. (3) in the main part) in the Neyman-Pearson approach to quantum hypothesis testing which, for $\alpha_0 \in [0,\, 1]$, is defined as

$$\begin{aligned} \beta^*_{\alpha_0}(\sigma,\, \rho) := \text{minimize } & \beta(M;\, \rho) \\ \text{s.t. } & \alpha(M;\, \sigma) \leq \alpha_0, \\ & 0 \leq M \leq \mathbb{1}_d. \end{aligned} \tag{6}$$

The following Lemmas lead to an explicit construction of Helstrom operators attaining a preassigned level of type-I error probability α_0 and which are optimizers of the SDP β^* (Eq. (3) in the main part). We first show that if a sequence of bounded Hermitian operators A_n converges in operator norm to a bounded Hermitian operator A from above (below), then the projection $\{A_n < 0\}$ ($\{A_n > 0\}$) converges to $\{A < 0\}$ ($\{A > 0\}$) in operator norm. This subsequently allows us to show that the function $t \mapsto \alpha(P_{t,+})$ is non-increasing and continuous from the right, and that $t \mapsto \alpha(P_{t,+} + P_{t,0})$ is non-increasing and continuous from the left. As a consequence, for $\alpha_0 \in [0,\, 1]$, the quantity

$$\tau(\alpha_0) := \inf\{t \geq 0\colon\; \alpha(P_{t,+}) \leq \alpha_0\} \tag{7}$$

is well defined. This implies the chain of inequalities

$$\alpha\left(P_{\tau(\alpha_0),+}\right) \leq \alpha_0 \leq \alpha\left(P_{\tau(\alpha_0),+} + P_{\tau(\alpha_0),0}\right). \tag{8}$$

Based on this, we construct a Helstrom operator $M_{\tau(\alpha_0)}$ according to

$$M_{\tau(\alpha_0)} := P_{\tau(\alpha_0),+} + q_0 P_{\tau(\alpha_0),0}, \;\; q_0 := \begin{cases} \frac{\alpha_0 - \alpha(P_{\tau(\alpha_0),+})}{\alpha(P_{\tau(\alpha_0),0})} & \text{if } \alpha\left(P_{\tau(\alpha_0),0}\right) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

which attains the preassigned type-I error probability α_0. We will show that these Helstrom operators are optimal for the SDP $\beta^*_{\alpha_0}(\sigma, \rho)$ in (6), so that

$$\beta^*_{\alpha_0}(\sigma, \rho) = \beta(M_{\tau(\alpha_0)}; \rho). \tag{10}$$

Lemma 1. *Denote by $\mathcal{B}(\mathcal{H})$ the space of bounded linear operators acting on the finite dimensional Hilbert space $\mathcal{H}$, $d := \dim(\mathcal{H}) < \infty$. Let $A \in \mathcal{B}(\mathcal{H})$ and $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{B}(\mathcal{H})$ be Hermitian operators and suppose that $\|A_n - A\|_{op} \xrightarrow{n\to\infty} 0$. Then, it holds that*

$$(i)\;\; A - A_n \leq 0 \;\Rightarrow\; \|\{A_n < 0\} - \{A < 0\}\|_{op} \xrightarrow{n\to\infty} 0, \tag{11}$$

$$(ii)\;\; A - A_n \geq 0 \;\Rightarrow\; \|\{A_n > 0\} - \{A > 0\}\|_{op} \xrightarrow{n\to\infty} 0. \tag{12}$$

Proof. We first show that convergence in operator norm implies that the eigenvalues of A_n converge towards the eigenvalues A. For a linear operator M let $\lambda_k(M)$ denote its k-th largest eigenvalue, $\lambda_1(M) \geq \ldots \geq \lambda_q(M)$, where $q \leq d$ is the number of distinct eigenvalues of M. By the minimax principle (e.g. [1], chapter 3), we can compute λ_k for any Hermitian operator M according to

$$\lambda_k(M) = \max_{\substack{V\subseteq\mathcal{H} \\ \dim(V)=k}} \min_{\substack{\psi\in V \\ \|\psi\|=1}} \langle\psi|M|\psi\rangle. \tag{13}$$

Now let $\varepsilon > 0$ and let $n \in \mathbb{N}$ large enough such that $\|A_n - A\|_{op} < \varepsilon$. Let $|\psi\rangle \in \mathcal{H}$ be a normalized state and note that by the Cauchy-Schwartz inequality we have

$$|\langle\psi|(A_n - A)|\psi\rangle| \leq \|(A_n - A)\psi\| \, \|\psi\| \leq \|A_n - A\|_{op} \|\psi\|^2 < \varepsilon \tag{14}$$

and thus

$$\langle\psi|A|\psi\rangle - \varepsilon < \langle\psi|A_n|\psi\rangle < \langle\psi|A|\psi\rangle + \varepsilon. \tag{15}$$

Hence, for any fixed $k \geq 1$ and any subspace $V \subset \mathcal{H}$ with $\dim(V) = k$, we have

$$\min_{\substack{\psi\in V \\ \|\psi\|=1}} \langle\psi|A|\psi\rangle - \varepsilon < \min_{\substack{\psi\in V \\ \|\psi\|=1}} \langle\psi|A_n|\psi\rangle < \min_{\substack{\psi\in V \\ \|\psi\|=1}} \langle\psi|A|\psi\rangle + \varepsilon \tag{16}$$

and thus, from (13), we see that

$$\lambda_k(A) - \varepsilon < \lambda_k(A_n) < \lambda_k(A) - \varepsilon \Rightarrow |\lambda_k(A) - \lambda_k(A_n)| < \varepsilon \tag{17}$$

and hence

$$\lambda_k(A_n) \xrightarrow{n\to\infty} \lambda_k(A), \qquad k = 1, \ldots, q. \tag{18}$$

Alternatively, this can be seen from Weyl's Perturbation Theorem (e.g. [1], ch. 3): namely, since A and A_n are Hermitian, it follows immediately from $|\lambda_k(A) - \lambda_k(A_n)| \leq \max_k |\lambda_k(A) - \lambda_k(A_n)| \leq \|A - A_n\|_{op}$ that eigenvalues converge provided that $\|A_n - A\|_{op} \to 0$. We will now make use of function theory and the resolvent formalism to show the convergence of the positive and negative eigenprojections. Let $M \in \mathcal{B}(\mathcal{H})$ be Hermitian, let $\sigma(M)$ denote the spectrum of M and, for $\lambda \in \mathbb{C} \setminus \sigma(M)$, let

$$R_\lambda(M) := (M - \lambda\mathbb{1})^{-1} = -\sum_{k=0}^{\infty} \lambda^{-(k+1)} M^k \tag{19}$$

be the resolvent of the operator M. The sum is the Neumann series and converges for $\lambda \in \mathbb{C} \setminus \sigma(M)$. Since M is Hermitian, we can write its spectral decomposition in terms of contour integrals over the resolvent

$$M = \sum_{k=1}^{q} \lambda_k(M) P_k \quad \text{with} \quad P_k = \frac{1}{2\pi \mathrm{i}} \oint_{(\gamma_k, -)} R_\lambda(M)\, d\lambda, \quad \text{and} \quad \sum_{k=1}^{q} P_k = \mathbb{1} \tag{20}$$

where P_k is the orthogonal projection onto the k-th eigenspace and the integration is to be understood element-wise. The symbol $(\gamma_k, -)$ indicates that the contour encircles $\lambda_k(M)$ once negatively, but does not encircle any other eigenvalue of M. We refer the reader to [2] for a detailed derivation.

We now show part (i) of the Lemma. For ease of notation, let λ_k denote the k-th eigenvalue of A and $\lambda_{k,n}$ the k-th eigenvalue of A_n. Since A_n and A are Hermitian operators, we can write the eigenprojections $\{A_n < 0\}$ and $\{A < 0\}$ in terms of the resolvent as

$$\{A < 0\} = \frac{1}{2\pi \mathrm{i}} \sum_{k:\, \lambda_k < 0} \oint_{(\gamma_k, -)} R_\lambda(A)\, d\lambda, \quad \{A_n < 0\} = \frac{1}{2\pi \mathrm{i}} \sum_{k:\, \lambda_{k,n} < 0} \oint_{(\gamma_{k,n}, -)} R_\lambda(A_n)\, d\lambda \tag{21}$$

where the symbols $(\gamma_k, -)$ and $(\gamma_{k,n}, -)$ indicate that the contours encircle only λ_k and $\lambda_{k,n}$ once negatively and no other eigenvalues of A and A_n respectively. Since by assumption $A_n \geq A$ and A_n, A are Hermitian, it follows from Weyl's Monotonicity Theorem that $\lambda_{k,n} \geq \lambda_k$. Let λ_K be the largest negative eigenvalue of A, that is $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_{K-1} \geq 0 > \lambda_K \geq \ldots \geq \lambda_q$. Note that if A is positive semidefinite, then so is A_n and the statement follows trivially from $\{A_n < 0\} = \{A < 0\} = 0$. Thus, without loss of generality, we can assume that at least one eigenvalue of A is negative. Since $\lambda_{k,n} \geq \lambda_k$, and in particular $\lambda_{K-1,n} \geq \lambda_{K-1} \geq 0$, there exists $N_0 \in \mathbb{N}$ large enough such that $\lambda_{K-1,n} \geq 0 > \lambda_{K,n}$ for all $n \geq N_0$. Let r_0 be the smallest distance between two eigenvalues of A

$$r_0 := \min_k |\lambda_k - \lambda_{k+1}| \tag{22}$$

and let $0 < \varepsilon < \frac{r_0}{2}$. Choose $N_1 \geq N_0$ large enough such that $\max_{k \geq K} |\lambda_{k,n} - \lambda_k| < \varepsilon/2$. Let $0 < \delta < \frac{r_0}{2} - \varepsilon$ and for $k \geq K$ let $B^k_{\delta+\varepsilon} := B_{\delta+\varepsilon}(\lambda_k)$ be the open ball of radius $\delta + \varepsilon$ centered at λ_k. Note that $\partial B^k_{\delta+\varepsilon}$ encircles both $\lambda_{k,n}$ and λ_k. Then, for $k \geq K$ and $n \geq N_1$, the mappings

$$\lambda \mapsto R_\lambda(A), \quad \lambda \in B^k_{\delta+\varepsilon} \setminus \{\lambda_k\}, \tag{23}$$

$$\lambda \mapsto R_\lambda(A_n), \quad \lambda \in B^k_{\delta+\varepsilon} \setminus \{\lambda_{k,n}\} \tag{24}$$

are holomorphic functions of λ and each has an isolated (simple) singularity at λ_k and $\lambda_{k,n}$ respectively. Let $\gamma_{k,n}$ be a contour around $\lambda_{k,n}$ encircling no other eigenvalue of A_n. Note that the contours $\gamma_{k,n}$ and $\partial B^k_{\delta+\varepsilon}$ are homotopic (in $B^k_{\delta+\varepsilon} \setminus \{\lambda_{k,n}\}$). Thus, for all $k \geq K$ and $n \geq N_1$, Cauchy's integral Theorem yields

$$\oint_{\gamma_{k,n}} R_\lambda(A_n)\, d\lambda = \oint_{\partial B^k_{\delta+\varepsilon}} R_\lambda(A_n)\, d\lambda. \tag{25}$$

With this, we see that for $n \geq N_1$

$$\{A_n < 0\} - \{A < 0\} = \frac{1}{2\pi \mathrm{i}} \sum_{k=K}^{q} \left(\oint_{\partial B^k_{\delta+\varepsilon}} R_\lambda(A)\, d\lambda - \oint_{\gamma_{k,n}} R_\lambda(A_n)\, d\lambda \right) \tag{26}$$

$$= \frac{1}{2\pi \mathrm{i}} \sum_{k=K}^{q} \left(\oint_{\partial B^k_{\delta+\varepsilon}} \left(R_\lambda(A) - R_\lambda(A_n) \right)\, d\lambda \right) \tag{27}$$

and thus, by the triangle inequality

$$\|\{A_n < 0\} - \{A < 0\}\|_{op} \leq \frac{1}{2\pi} \sum_{k=K}^{q} \left\| \oint_{\partial B^k_{\delta+\varepsilon}} \left(R_\lambda(A) - R_\lambda(A_n) \right)\, d\lambda \right\|_{op} \tag{28}$$

$$\leq \frac{1}{2\pi} \sum_{k=K}^{q} \sup_{\lambda \in \partial B^k_{\delta+\varepsilon}} \|R_\lambda(A) - R_\lambda(A_n)\|_{op} \cdot 2\pi \cdot (\delta + \varepsilon) \tag{29}$$

$$\leq (q - (K-1)) \cdot (\delta + \varepsilon) \cdot \max_{k \geq K} \left(\sup_{\lambda \in \partial B^k_{\delta+\varepsilon}} \|R_\lambda(A) - R_\lambda(A_n)\|_{op} \right). \tag{30}$$

Furthermore, for any k and $\lambda \in \partial B^k_{\delta+\varepsilon}$, the second resolvent identity yields

$$\begin{aligned}\|R_\lambda(A) - R_\lambda(A_n)\|_{op} &= \|R_\lambda(A)(A - A_n)R_\lambda(A)\|_{op} \\ &\leq \|A - A_n\|_{op} \cdot \|R_\lambda(A)\|_{op} \|R_\lambda(A_n)\|_{op} .\end{aligned} \tag{31}$$

We now show that the supremum over $\lambda \in \partial B^k_{\delta+\varepsilon}$ in the right hand side of (31) is bounded. Since both A and A_n are Hermitian, it follows that their resolvent is normal and bounded for $\lambda \in \mathbb{C} \setminus \sigma(A_n)$ and $\lambda \in \mathbb{C} \setminus \sigma(A)$ respectively. The operator norm is thus given by the spectral radius,

$$\|R_\lambda(A)\|_{op} = \max_k |\lambda_k(R_\lambda(A))| \quad \text{and} \quad \|R_\lambda(A_n)\|_{op} = \max_k |\lambda_k(R_\lambda(A_n))| . \tag{32}$$

Note that the eigenvalues of $R_\lambda(A)$ are given by $(\lambda_k(A) - \lambda)^{-1}$. To see this, let $\lambda \in \mathbb{C} \setminus \sigma(A)$ and consider

$$\det\left(R_\lambda(A) - \mu\mathbb{1}\right) = \det\left((A - \lambda\mathbb{1})^{-1}\left(\mathbb{1} - (A - \lambda\mathbb{1})\mu\mathbb{1}\right)\right) \tag{33}$$

$$\propto \det\left(\mathbb{1} - (A - \lambda\mathbb{1})\mu\mathbb{1}\right) = (-\mu)^m \det\left(A - (\mu^{-1} + \lambda)\mathbb{1}\right). \tag{34}$$

Since $\det(R_\lambda(A)) \neq 0$ it follows that $\mu = 0$ can not be an eigenvalue. Thus, eigenvalues of $R_\lambda(A)$ satisfy

$$\frac{1}{\mu} + \lambda = \lambda_k(A) \Rightarrow \mu = \frac{1}{\lambda_k(A) - \lambda}. \tag{35}$$

The same reasoning yields an expression for eigenvalues of $R_\lambda(A_n)$. Thus

$$\|R_\lambda(A)\|_{op} = \max_k \frac{1}{|\lambda_k(A) - \lambda|} \quad \text{and} \quad \|R_\lambda(A_n)\|_{op} = \max_k \frac{1}{|\lambda_k(A_n) - \lambda|}. \tag{36}$$

Note that, by the definition of δ, for $\lambda \in \partial B^k_{\delta+\varepsilon}$, the eigenvalue of A which is nearest to λ is given by $\lambda_k(A)$. Since this is exactly the center of the ball $B_{\delta+\varepsilon}$, it follows that

$$\sup_{\lambda \in \partial B^k_{\delta+\varepsilon}} \|R_\lambda(A)\|_{op} = \frac{1}{\delta + \varepsilon} < \infty. \tag{37}$$

Similarly, for $\lambda \in \partial B^k_{\delta+\varepsilon}$, the eigenvalue of A_n which is nearest to λ is given $\lambda_k(A_n)$ since n was chosen large enough such that $|\lambda_k(A_n) - \lambda_k(A)| < \varepsilon$ and $\varepsilon < r_0/2$. Since $\delta < \frac{r_0}{2} - \varepsilon$, it follows that the smallest distance from $\partial B^k_{\delta+\varepsilon}$ to $\lambda_k(A_n)$ is exactly δ and thus

$$\sup_{\lambda \in \partial B^k_{\delta+\varepsilon}} \|R_\lambda(A_n)\|_{op} = \frac{1}{\delta} < \infty. \tag{38}$$

Hence, we find that the RHS in (31) is bounded by $\frac{\|A_n - A\|_{op}}{(\delta+\varepsilon)\delta}$ for $\lambda \in \partial B^k_{\delta+\varepsilon}$. Finally, this yields

$$\|\{A_n < 0\} - \{A < 0\}\|_{op} \leq (\delta + \varepsilon)(q - (K - 1)) \max_{k \geq K} \left(\sup_{\lambda \in \partial B^k_{\delta+\varepsilon}} \|R_\lambda(A) - R_\lambda(A_n)\|_{op} \right) \tag{39}$$

$$\leq (\delta + \varepsilon)(q - (K - 1)) \|A_n - A\|_{op} \frac{1}{\delta + \varepsilon} \frac{1}{\delta} \tag{40}$$

$$= \frac{q - (K - 1)}{\delta} \|A_n - A\|_{op} \xrightarrow{n \to \infty} 0. \tag{41}$$

In an analogous way we can show that

$$\|\{A_n > 0\} - \{A > 0\}\|_{op} \leq \frac{R}{\delta} \|A_n - A\|_{op} \xrightarrow{n \to \infty} 0 \tag{42}$$

where R denotes the index of the smallest positive eigenvalue of A. This concludes the proof. □

Lemma 2. *The functions $t \mapsto \alpha(P_{t,+})$ and $t \mapsto \alpha(P_{t,+} + P_{t,0})$ are non-increasing.*

Proof. Let $0 \leq s < t$. We need to show that $\mathrm{Tr}\,[\sigma P_{s,+}] \geq \mathrm{Tr}\,[\sigma P_{t,+}]$ and $\mathrm{Tr}\,[\sigma(P_{s,+} + P_{s,0})] \geq \mathrm{Tr}\,[\sigma(P_{t,+} + P_{t,0})]$. Note that for any Hermitian operator A on and for any Hermitian operator $0 \leq T \leq \mathbb{1}$ we have

$$\mathrm{Tr}\,[A\{A \geq 0\}] = \mathrm{Tr}\,[A\{A > 0\}] \geq \mathrm{Tr}\,[A \cdot T] \tag{43}$$

We first show $\mathrm{Tr}\,[\sigma P_{s,+}] \geq \mathrm{Tr}\,[\sigma P_{t,+}]$. Write $T_s := \rho - s\sigma$ and $T_t := \rho - t\sigma$ and note that the eigenprojections are Hermitian and satisfy

$$0 \leq \{T_s > 0\} \leq \mathbb{1}, \text{ and } \quad 0 \leq \{T_t > 0\} \leq \mathbb{1} \tag{44}$$

It follows that

$$\mathrm{Tr}\,[T_t\{T_t > 0\}] \geq \mathrm{Tr}\,[T_t\{T_s > 0\}] \tag{45}$$

and similarly

$$\mathrm{Tr}\,[T_s\{T_s > 0\}] \geq \mathrm{Tr}\,[T_s\{T_t > 0\}]\,. \tag{46}$$

Combining (45) and (46) yields

$$t \cdot (\mathrm{Tr}\,[\sigma\{T_s > 0\}] - \mathrm{Tr}\,[\sigma\{T_t > 0\}]) \geq s \cdot (\mathrm{Tr}\,[\sigma\{T_s > 0\}] - \mathrm{Tr}\,[\sigma\{T_t > 0\}])\,. \tag{47}$$

Since $s < t$ it follows that $\mathrm{Tr}\,[\sigma\{T_s > 0\}] \geq \mathrm{Tr}\,[\sigma\{T_t > 0\}]$ and thus $\alpha(P_{s,+}) \geq \alpha(P_{t,+})$. The other statement follows analogously by replacing $\{T_t > 0\}$ and $\{T_s > 0\}$ by $\{T_t \geq 0\}$ and $\{T_s \geq 0\}$. This concludes the proof. □

Lemma 3. *The function $t \mapsto \alpha(P_{t,+})$ is continuous from the right.*

Proof. Let $t \geq 0$ and let $\{t_n\}_{n\in\mathbb{N}} \subseteq [0, \infty)$ be a sequence such that $t_n \downarrow t$ (i.e. t_n converges to t from above). We show that $\lim_{n\to\infty} \alpha(P_{t_n,+}) = \alpha(P_{t,+})$. Define the operators

$$A_n := \rho - t_n\sigma, \;\; A := \rho - t\sigma \tag{48}$$

and note that

$$\|A_n - A\|_{op} = |t_n - t| \cdot \|\sigma\|_{op} \xrightarrow{n\to\infty} 0. \tag{49}$$

Since, in addition, both A_n and A are Hermitian and $A - A_n = (t_n - t)\sigma \geq 0$, it follows from the second part of Lemma 1 that

$$\|\{A_n > 0\} - \{A > 0\}\|_{op} \xrightarrow{n\to\infty} 0 \tag{50}$$

and thus

$$\alpha(P_{t_n,+}) = \mathrm{Tr}\,[\sigma\{A_n > 0\}] \xrightarrow{n\to\infty} \mathrm{Tr}\,[\sigma\{A > 0\}] = \alpha(P_{t,+}) \tag{51}$$

since operator norm convergence implies convergence in the weak operator topology. This concludes the proof. □

Lemma 4. *The function $t \mapsto \alpha(P_{t,+} + P_{t,0})$ is continuous from the left.*

Proof. Let $\{t_n\}_{n\in\mathbb{N}} \subset [0, \infty)$ be a sequence of non-negative real numbers such that $t_n \uparrow t$ (i.e. t_n converges to t from below). Let A_n and A be the Hermitian operators defined by

$$A_n := \rho - t_n\sigma, \;\; A := \rho - t\sigma \tag{52}$$

and note that $A - A_n = (t_n - t)\sigma \leq 0$ and $\|A_n - A\|_{op} \to 0$ as $n \to \infty$. It follows from the first part of Lemma 1 that

$$\|\{A_n < 0\} - \{A < 0\}\|_{op} \xrightarrow{n\to\infty} 0 \tag{53}$$

and thus, since operator norm convergence implies convergence in the weak operator topology, we have

$$\alpha(P_{t_n,+} + P_{t_n,0}) = \mathrm{Tr}\,[\sigma(\mathbb{1} - \{A_n < 0\})] \xrightarrow{n\to\infty} \mathrm{Tr}\,[\sigma(\mathbb{1} - \{A < 0\})] = \alpha(P_{t,+} + P_{t,0}). \tag{54}$$

This concludes the proof. □

Lemma 5. *For $\alpha_0 \in [0,\, 1]$ we have the chain of inequalities*

$$\alpha\left(P_{\tau(\alpha_0),+}\right) \leq \alpha_0 \leq \alpha\left(P_{\tau(\alpha_0),+} + P_{\tau(\alpha_0),0}\right). \tag{55}$$

Proof. Recall that $\tau(\alpha_0) := \inf\{t \geq 0 \colon \alpha(P_{t,+}) \leq \alpha_0\}$. Since, by Lemmas 2 and 3 the function $t \mapsto \alpha(P_{t,+})$ is non-decreasing and right-continuous, the left hand side of the inequality follows directly from the definition of $\tau(\alpha_0)$.

We now show the right hand side of the inequality. Note that if $t := \tau(\alpha_0) = 0$, then $P_{t,-} = \{\rho < 0\} = 0$ and thus $P_{t,+} + P_{t,0} = \mathbb{1}$ and $\alpha(P_{t,+} + P_{t,0}) = \operatorname{Tr}[\sigma] = 1 \geq \alpha_0$. Suppose that $\tau(\alpha_0) > 0$ and let $\{t_n\}_{n\in\mathbb{N}} \subset [0,\infty)$ be a sequence of non-negative real numbers such that $t_n \uparrow \tau(\alpha_0)$. Note that, since $t_n < \tau(\alpha_0)$ for all n and $t \mapsto \alpha(P_{t,+})$ is non-increasing and right-continuous by Lemmas 2 and 3, we have

$$\operatorname{Tr}[\sigma P_{t_n,+}] = \alpha(P_{t_n,+}) > \alpha_0 \tag{56}$$

by definition of $\tau(\alpha_0)$. Furthermore, since the projection $P_{t_n,0}$ is positive semidefinite, it follows that

$$\alpha(P_{t_n,+} + P_{t_n,0}) = \operatorname{Tr}[\sigma(P_{t_n,+} + P_{t_n,0})] \geq \operatorname{Tr}[\sigma P_{t_n,+}] = \alpha(P_{t_n,+}) > \alpha_0. \tag{57}$$

The Lemma now follows since by Lemma 4 the function $t \mapsto \alpha(P_{t,+} + P_{t,0})$ is continuous from the left and hence

$$\alpha(P_{\tau(\alpha_0),+} + P_{\tau(\alpha_0),0}) = \lim_{n\to\infty} \alpha(P_{t_n,+} + P_{t_n,0}) \geq \alpha_0 \tag{58}$$

which concludes the proof. □

Supplementary Note 2. CONSTRUCTION OF HELSTROM OPERATORS AND OPTIMALITY

Given $\alpha_0 \in [0,\, 1]$, it follows from Lemma 5 that firstly $\alpha_0 - \alpha\left(P_{\tau(\alpha_0),+}\right) \geq 0$, and secondly, since $\alpha_0 - \alpha\left(P_{\tau(\alpha_0),+}\right) \leq \alpha\left(P_{\tau(\alpha_0),+} + P_{\tau(\alpha_0),0}\right) - \alpha\left(P_{\tau(\alpha_0),+}\right)$ we have that

$$\frac{\alpha_0 - \alpha\left(P_{\tau(\alpha_0),+}\right)}{\alpha\left(P_{\tau(\alpha_0),0}\right)} \in [0,\, 1] \tag{59}$$

if $\alpha\left(P_{\tau(\alpha_0),0}\right) \neq 0$. It follows that, for any $\alpha_0 \in [0,\, 1]$, a Helstrom operator with type-I error probability α_0 is given by

$$M_{\tau(\alpha_0)} := P_{\tau(\alpha_0),+} + q_0 P_{\tau(\alpha_0),0}, \quad \text{where} \quad q_0 := \begin{cases} \frac{\alpha_0 - \alpha(P_{\tau(\alpha_0),+})}{\alpha(P_{\tau(\alpha_0),0})} & \text{if } \alpha\left(P_{\tau(\alpha_0),0}\right) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{60}$$

To see that $M_{\tau(\alpha_0)}$ indeed has type-I error probability α_0, note that if $q_0 \neq 0$ then

$$\alpha(M_{\tau(\alpha_0)}) = \operatorname{Tr}\left[\sigma M_{\tau(\alpha_0)}\right] = \operatorname{Tr}\left[\sigma P_{\tau(\alpha_0),+}\right] + q_0 \operatorname{Tr}\left[\sigma P_{\tau(\alpha_0),0}\right] = \alpha_0. \tag{61}$$

If on the other hand $q_0 = 0$, then, by Lemma 5, we have $\alpha_0 = \alpha(P_{\tau(\alpha_0),+}) = \alpha(M_{\tau(\alpha_0)})$. The following lemma shows optimality of the Helstrom operators, i.e. equation (10):

Lemma 6. *Let $t \geq 0$ and let $M_t := P_{t,+} + X_t$ for $0 \leq X_t \leq P_{t,0}$ be a Helstrom operator. Then, for any quantum hypothesis test $0 \leq M \leq \mathbb{1}$ for testing the null σ against the alternative ρ, the following implications hold*

(i) $\alpha(M) \leq \alpha(M_t) \;\Rightarrow\; \beta(M) \geq \beta(M_t)$

(ii) $\alpha(M) \geq 1 - \alpha(M_t) \;\Rightarrow\; 1 - \beta(M) \geq \beta(M_t)$

Proof. Let $\sum_i \lambda_i P_i$ be the spectral decomposition of the operator $\rho - t\sigma$ with orthogonal projections $\{P_i\}_i$ and associated eigenvalues $\{\lambda_i\}_i$. Recall that

$$P_{t,+} := \sum_{i\colon \lambda_i > 0} P_i \;\; P_{t,-} := \sum_{i\colon \lambda_i < 0} P_i, \;\; P_{t,0} := \mathbb{1} - P_{t,+} - P_{t,-}. \tag{62}$$

We notice that for any $0 \leq X_t \leq P_{t,0}$ we have

$$\operatorname{Tr}[(\rho - t\sigma)X_t] = \operatorname{Tr}[(\rho - t\sigma)P_{t,+}X_t] + \operatorname{Tr}[(\rho - t\sigma)P_{t,-}X_t] \leq \operatorname{Tr}[(\rho - t\sigma)P_{t,+}P_{t,0}] = 0 \tag{63}$$

and

$$\mathrm{Tr}\left[(\rho - t\sigma)X_t\right] = \mathrm{Tr}\left[(\rho - t\sigma)P_{t,+}X_t\right] + \mathrm{Tr}\left[(\rho - t\sigma)P_{t,-}X_t\right] \geq \mathrm{Tr}\left[(\rho - t\sigma)P_{t,-}P_{t,0}\right] = 0 \tag{64}$$

and thus $\mathrm{Tr}\left[(\rho - t\sigma)P_{t,0}\right] = \mathrm{Tr}\left[(\rho - t\sigma)X_t\right] = 0$. For the sequel, let $\bar{M}_t := \mathbb{1} - M_t$ and $\bar{M} := \mathbb{1} - M$.

We first show part (i) of the statement. Multiplying with the identity yields

$$M - M_t = (\bar{M}_t + M_t)M - M_t(\bar{M} + M) = \bar{M}_t M - M_t\bar{M} \tag{65}$$

and adding zero yields

$$\rho(M - M_t) = (\rho - t\sigma)(M - M_t) + t\sigma(M - M_t) \tag{66}$$

$$= (\rho - t\sigma)(\bar{M}_t M - M_t\bar{M}) + t\sigma(\bar{M}_t M - M_t\bar{M}). \tag{67}$$

We need to show that $\beta(M_t) - \beta(M) = \mathrm{Tr}\left[\rho(M - M_t)\right] \leq 0$. Notice that

$$\mathrm{Tr}\left[(\rho - t\sigma)\bar{M}_t M\right] = -\mathrm{Tr}\left[(\rho - t\sigma)_- M\right] \leq 0 \tag{68}$$

and similarly

$$\mathrm{Tr}\left[(\rho - t\sigma)M_t\bar{M}\right] = \mathrm{Tr}\left[(\rho - t\sigma)_+\bar{M}\right] \geq 0 \tag{69}$$

where the inequalities follow from $\mathbb{1} \geq M \geq 0$. Finally, we see that

$$\mathrm{Tr}\left[\rho(M - M_t)\right] = \mathrm{Tr}\left[(\rho - t\sigma)(\bar{M}_t M - M_t\bar{M})\right] + t\cdot\mathrm{Tr}\left[\sigma(\bar{M}_t M - M_t\bar{M})\right] \tag{70}$$

$$\leq t\cdot\mathrm{Tr}\left[\sigma(\bar{M}_t M - M_t\bar{M})\right] \tag{71}$$

$$= t\cdot\mathrm{Tr}\left[\sigma(M - M_t)\right] \tag{72}$$

$$= t\cdot(\alpha(M) - \alpha(M_t)) \leq 0 \tag{73}$$

where the last inequality follows from the assumption and $t \geq 0$.

Part (ii) now follows directly from part (i) by noting that $0 \leq M' := \mathbb{1} - M \leq \mathbb{1}$ and

$$\alpha(M) \geq 1 - \alpha(M_t) \Rightarrow \alpha(M') \leq \alpha(M_t) \overset{(i)}{\Longrightarrow} \beta(M_t) \leq \beta(M') = 1 - \beta(M). \tag{74}$$

This concludes the proof. □

Supplementary Note 3. PROOF OF LEMMA 1

Lemma 1 (restated)**.** *Let $|\psi_\sigma\rangle$, $|\psi_\rho\rangle \in \mathcal{H}$ and let $\mathcal{A}$ be a quantum classifier. Suppose that for $k_A \in \mathcal{C}$ and p_{A}, $p_{\mathrm{B}} \in [0, 1]$, we have $k_A = \mathcal{A}(\psi_\sigma)$ and suppose that the score function $\mathbf{y}$ satisfies (9). Then, it is guaranteed that $\mathcal{A}(\psi_\rho) = \mathcal{A}(\psi_\sigma)$ for any ψ_ρ with*

$$|\langle\psi_\sigma|\psi_\rho\rangle|^2 > \frac{1}{2}\left(1 + \sqrt{g(p_{\mathrm{A}}, p_{\mathrm{B}})}\right), \tag{75}$$

where the function g is given by

$$g(p_{\mathrm{A}}, p_{\mathrm{B}}) = 1 - p_{\mathrm{B}} - p_{\mathrm{A}}(1 - 2p_{\mathrm{B}}) + 2\sqrt{p_{\mathrm{A}}p_{\mathrm{B}}(1 - p_{\mathrm{A}})(1 - p_{\mathrm{B}})}. \tag{76}$$

This condition is equivalent to (10) and is hence both sufficient and necessary whenever $p_{\mathrm{A}} + p_{\mathrm{B}} = 1$.

Proof. In order to prove the lemma, we show that the fidelity bound in eq. (12) is equivalent to the SDP robustness condition (10) from Theorem 1 expressed in terms of type-II error probabilities. For that purpose, we first derive an expression for the Helstrom operators in terms of the squared overlap between $|\psi_\rho\rangle$ and $|\psi_\sigma\rangle$ and subsequently solve (10) for $|\langle\psi_\sigma|\psi_\rho\rangle|^2$. For the sequel, let $\sigma = |\psi_\sigma\rangle\langle\psi_\sigma|$ and $\rho = |\psi_\rho\rangle\langle\psi_\rho|$. Recall that a Helstrom operator with type-I error probability α_0 takes the form (60) which reads

$$M_{\tau(\alpha_0)} := P_{\tau(\alpha_0),+} + q_0 P_{\tau(\alpha_0),0},\;\; q_0 := \begin{cases} \frac{\alpha_0 - \alpha(P_{\tau(\alpha_0),+})}{\alpha(P_{\tau(\alpha_0),0})} & \text{if } \alpha\left(P_{\tau(\alpha_0),0}\right) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{77}$$

where $\tau(\alpha_0) := \inf\{t \geq 0\colon\ \alpha(P_{t,+}) \leq \alpha_0\}$. We now proceed as follows. We first compute the spectral decomposition of the operator $\rho - t\sigma$ as a function of t. With this, we derive an expression for $\alpha(P_{t,+})$ and subsequently compute $\tau(\alpha_0)$. This yields an expression for the Helstrom operators with type-I error probabilities $1 - p_{\mathrm{A}}$ and p_{B} which can then be used to solve inequality (10) for the fidelity. We thus start by solving the eigenvalue problem

$$(\rho - t\sigma)|\eta\rangle = \eta|\eta\rangle. \tag{78}$$

Since σ and ρ are both pure states, the operator $\rho - t\sigma$ is of rank at most 2. It follows that there are at most two states $|\eta_0\rangle$ and $|\eta_1\rangle$ satisfying (78) with nonzero eigenvalues and, in addition, they are linear combinations of $|\psi_\sigma\rangle$ and $|\psi_\rho\rangle$

$$|\eta_k\rangle = z_{k,\sigma}|\psi_\sigma\rangle + z_{k,\rho}|\psi_\rho\rangle, \qquad k = 0,\, 1 \tag{79}$$

with constants $z_{k,\sigma}$ and $z_{k,\rho}$ that are to be determined. Substituting this into (78) yields the system of equations

$$\begin{aligned} z_{k,\rho} + \gamma z_{k,\sigma} &= \eta_k z_{k,\rho} \\ -t\gamma^\dagger z_{k,\rho} - t z_{k,\sigma} &= \eta_k z_{k,\sigma} \end{aligned} \tag{80}$$

where $\gamma := \langle\psi_\rho|\psi_\sigma\rangle$ is the overlap between states ρ and σ. The two eigenvalues η_k for which these equations possess nonzero solutions are given by

$$\begin{aligned} \eta_0 = \frac{1}{2}(1-t) + R, \qquad \eta_1 = \frac{1}{2}(1-t) - R, \\ R = \sqrt{\frac{1}{4}(1-t)^2 + t\left(1 - |\gamma|^2\right)} \end{aligned} \tag{81}$$

with $\eta_0 > 0$ and $\eta_1 \leq 0$. With the condition $\langle\eta_k|\eta_k\rangle = 1$, the coefficients $z_{k,\sigma}$ and $z_{k,\rho}$ are then determined as

$$\begin{aligned} z_{k,\rho} = -\gamma A_k, \qquad z_{k,\sigma} = (1-\eta_k)A_k, \\ |A_k|^{-2} = 2R\left|\eta_k - 1 + |\gamma|^2\right|. \end{aligned} \tag{82}$$

Recall that $P_{t,+} = \sum_{k\colon\ \eta_k > 0} P_k$ and hence $P_{t,+} = |\eta_0\rangle\langle\eta_0|$. We thus obtain the expression

$$\alpha(P_{t,+}) = \mathrm{Tr}\left[\sigma P_{t,+}\right] = |\langle\eta_0|\psi_\sigma\rangle|^2 \tag{83}$$

$$= |z_{0,\sigma} + z_{0,\rho}\gamma^*|^2 \tag{84}$$

$$= |A_0|^2\left|1 - \eta_0 - |\gamma|^2\right|^2 \tag{85}$$

$$= \frac{\eta_0 - 1 + |\gamma|^2}{2R} \tag{86}$$

Substituting in the expressions for η_0 and R yields

$$\alpha(P_{t,+}) = \frac{\frac{1}{2}(1-t) + \sqrt{\frac{1}{4}(1-t)^2 + t\left(1-|\gamma|^2\right)} - 1 + |\gamma|^2}{2\sqrt{\frac{1}{4}(1-t)^2 + t\left(1-|\gamma|^2\right)}} = \frac{1}{2}\left(1 - \frac{1 + t - 2|\gamma|^2}{\sqrt{(1+t)^2 - 4t|\gamma|^2}}\right) \tag{87}$$

The next step is to compute $t_A = \tau(1 - p_{\mathrm{A}})$ and $t_B = \tau(p_{\mathrm{B}})$. By Lemma 2, the function $t \mapsto g(t) := \alpha(P_{t,+})$ is non-increasing and thus attains its maximum at $t = 0$

$$g(0) = \alpha(P_{0,+}) = |\gamma|^2 \tag{88}$$

Furthermore, note that the only real non-negative discontinuity of g is located at $t = 1$ in the case where $|\gamma|^2 = 1$. Since this corresponds to two identical states we exclude this case in the following and assume $|\gamma| \in [0, 1)$. Notice that, if $\alpha_0 \geq |\gamma|^2$, then we have that $\tau(\alpha_0) = 0$. Otherwise, if $\alpha_0 < |\gamma|^2$, then we have that $\alpha(P_{t,+}) = \alpha_0$ if $t \geq 0$ is the non-negative root of the polynomial

$$t \mapsto t^2 + 2t(1 - 2|\gamma|^2) + \left(1 - \frac{|\gamma|^2(1 - |\gamma|^2)}{\alpha_0(1-\alpha_0)}\right) \tag{89}$$

which is calculated as

$$t = 2|\gamma|^2 - 1 - (2\alpha_0 - 1)\sqrt{\frac{|\gamma|^2(1-|\gamma|^2)}{\alpha_0(1-\alpha_0)}}. \tag{90}$$

Thus, in summary, we find that $\tau(\alpha_0)$ is given by

$$\tau(\alpha_0) = \begin{cases} 2|\gamma|^2 - 1 - (2\alpha_0 - 1)\sqrt{\frac{|\gamma|^2(1-|\gamma|^2)}{\alpha_0(1-\alpha_0)}} & \text{if } \alpha_0 < |\gamma|^2, \\ 0 & \text{if } \alpha_0 \geq |\gamma|^2. \end{cases} \tag{91}$$

First, we notice that if $|\gamma|^2 \leq \min\{p_\mathrm{B},\, 1-p_\mathrm{A}\}$, then we have $t_A = \tau(1-p_\mathrm{A}) = 0$ and $t_B = \tau(p_\mathrm{B}) = 0$ and hence the robustness condition (10) cannot be satisfied. In the case where $\min\{p_\mathrm{B},\, 1-p_\mathrm{A}\} < |\gamma|^2 \leq \max\{p_\mathrm{B},\, 1-p_\mathrm{A}\}$, then either $t_A = 0$ or $t_B = 0$. Without loss of generality, suppose that $t_A = 0$. Then the Helstrom operator takes the form $M_{\tau(1-p_\mathrm{A})} = \{\rho > 0\} + q_A(\mathbb{1} - \{\rho > 0\} - \{\rho < 0\})$ and thus

$$\beta(M_{\tau(1-p_\mathrm{A})}) = 1 - \mathrm{Tr}\left[\rho\left(\{\rho > 0\} + q_A(\mathbb{1} - \{\rho > 0\} - \{\rho < 0\})\right)\right] = 0. \tag{92}$$

In particular, it follows that the robustness condition $\beta^*_{1-p_\mathrm{A}}(\sigma,\, \rho) + \beta^*_{p_\mathrm{B}}(\sigma,\, \rho) > 1$ cannot be satisfied. The same follows in the case where $t_B = 0$. Finally, if $|\gamma|^2 > \max\{p_\mathrm{B},\, 1-p_\mathrm{A}\}$, then we have that $t_A > 0$ and $t_B > 0$. We notice that $\alpha(P_{t_A,+}) = 1-p_\mathrm{A}$ and $\alpha(P_{t_B,+}) = p_\mathrm{B}$ and thus $M_{\tau(1-p_\mathrm{A})} = P_{t_A,+}$ and $M_{\tau(p_\mathrm{B})} = P_{t_B,+}$. Computing the type-II error for $M_{\tau(1-p_\mathrm{A})}$ yields

$$\beta(M_{\tau(1-p_\mathrm{A})}) = 1 - \mathrm{Tr}\left[\rho P_{t_A,+}\right] = 1 - |\langle\eta_0|\psi_\rho\rangle|^2 = 1 - |A_0|^2|\gamma|^2|\eta_0|^2 \tag{93}$$

$$= |\gamma|^2(2p_\mathrm{A} - 1) + (1-p_\mathrm{A})\left(1 - 2p_\mathrm{A}\sqrt{\frac{|\gamma|^2(1-|\gamma|^2)}{p_\mathrm{A}(1-p_\mathrm{A})}}\right) \tag{94}$$

where $|\eta_0\rangle$ corresponds to the eigenvector associated with the eigenvalue η_0 at $t = t_A$. Similarly, computing the type-II error for $M_{\tau(p_\mathrm{B})}$ yields

$$\beta(M_{\tau(p_\mathrm{B})}) = 1 - \mathrm{Tr}\left[\rho P_{t_B,+}\right] = 1 - |\langle\eta_0|\psi_\rho\rangle|^2 = 1 - |A_0|^2|\gamma|^2|\eta_0|^2 \tag{95}$$

$$= |\gamma|^2(1 - 2p_\mathrm{B}) + p_\mathrm{B}\left(1 - 2(1-p_\mathrm{B})\sqrt{\frac{|\gamma|^2(1-|\gamma|^2)}{p_\mathrm{B}(1-p_\mathrm{B})}}\right) \tag{96}$$

where $|\eta_0\rangle$ corresponds to the eigenvector associated with the eigenvalue η_0 at $t = t_B$. With these expressions, it follows that the robustness condition (10), i.e. $\beta(M_{\tau(1-p_\mathrm{A})}) + \beta(M_{\tau(p_\mathrm{B})}) > 1$, is equivalent to

$$|\langle\psi_\sigma|\psi_\rho\rangle|^2 > \frac{1}{2}\left(1 + \sqrt{1 - p_\mathrm{B} - p_\mathrm{A}(1-2p_\mathrm{B}) + 2\sqrt{p_\mathrm{A}p_\mathrm{B}(1-p_\mathrm{A})(1-p_\mathrm{B})}}\right) \tag{97}$$

what concludes the proof. □

Supplementary Note 4. PROOF OF COROLLARY 1

Corollary 1 (restated)**.** *Let $\sigma,\, \rho \in \mathcal{S}(\mathcal{H})$ and suppose that $\sigma = |\psi_\sigma\rangle\langle\psi_\sigma|$ is pure. Let $\mathcal{A}$ be a quantum classifier and suppose that for $k_A \in \mathcal{C}$ and $p_\mathrm{A},\, p_\mathrm{B} \in [0,\, 1]$, we have $k_A = \mathcal{A}(\sigma)$ and suppose that the score function* $\mathbf{y}$ *satisfies* (9)*. Then, it is guaranteed that $\mathcal{A}(\rho) = \mathcal{A}(\sigma)$ for any ρ with*

$$T(\rho,\, \sigma) < \delta(p_\mathrm{A},\, p_\mathrm{B})\left(1 - \sqrt{1 - \delta(p_\mathrm{A},\, p_\mathrm{B})^2}\right) \tag{98}$$

where $\delta(p_\mathrm{A},\, p_\mathrm{B}) = [\frac{1}{2}(1 - g(p_\mathrm{A}, p_\mathrm{B}))]^{\frac{1}{2}}$.

Proof. We denote the convex hull enclosed by the set of robust pure states as $\mathcal{C} := \mathrm{Conv}(\{|\psi\rangle\langle\psi| : \||\psi\rangle\langle\psi| - \sigma\|_1 < \delta(P_A, P_B)\})$. Observe that any convex mixture $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$ with $\sum_i p_i = 1$ of any sets of robust pure states $\{|\psi_i\rangle\langle\psi_i|\} \in \mathcal{C}$ must also be robust. Thus it suffices to prove condition (21) implies $\rho \in \mathcal{C}$. Note that the boundary consisting of non-extreme points (which correspond to mixed-states) of $\mathcal{C}$ interfaces with the set $\mathcal{C}^* = \mathrm{Conv}(\{|\psi^*\rangle\langle\psi^*| : \||\psi^*\rangle\langle\psi^*| - \sigma\|_1 = \delta(P_A, P_B)\})$. Thus, it suffices to compute the shortest distance r from σ to $\mathcal{C}^*$, such that $r = \min_{\rho^*} \|\rho^* - \sigma\|_1$ where $\rho^* \in \mathcal{C}^*$, then $\|\rho - \sigma\|_1 < r$ guarantees robustness. Further note that for every ρ^*, $\exists D_\sigma \rho^* D_\sigma^\dagger \in \mathcal{C}^*$, where $D_\sigma = 2\sigma - \mathbb{1}$, such that $\|\rho^* - \sigma\|_1 = \|D_\sigma \rho^* D_\sigma^\dagger - \sigma\|_1$, and $\|p_1\rho^* + p_2 D_\sigma \rho^* D_\sigma^\dagger - \sigma\|_1 < \|\rho^* - \sigma\|_1$ for $p_1 + p_2 = 1$, and $p_1 \neq 0$, $p_2 \neq 0$. Therefore to minimise the distance to σ, it suffices to require $\rho^* = D_\sigma \rho^* D_\sigma^\dagger$, a valid of which is $\rho^* = \frac{1}{2}(|\psi^*\rangle\langle\psi^*| + D_\sigma|\psi^*\rangle\langle\psi^*|D_\sigma^\dagger)$. As such we have

$$\begin{aligned} r &= \|\sigma - \frac{1}{2}(|\psi^*\rangle\langle\psi^*| + D_\sigma|\psi^*\rangle\langle\psi^*|D_\sigma^\dagger)\|_1 \\ &= \||\psi^*\rangle\langle\psi^*| - \sigma + 2|\langle\psi_\sigma|\psi^*\rangle|^2\sigma - \langle\psi_\sigma|\psi^*\rangle|\psi_\sigma\rangle\langle\psi| - \langle\psi|\psi_\sigma\rangle|\psi\rangle\langle\psi_\sigma|\|_1 \end{aligned} \tag{99}$$

Note that we have $\||\psi^*\rangle\langle\psi^*| - \sigma\|_1 = \delta(P_A, P_B)$ by definition and that

$$\begin{aligned} \|2|\langle\psi_\sigma|\psi^*\rangle|^2\sigma - \langle\psi_\sigma|\psi^*\rangle|\psi_\sigma\rangle\langle\psi| - \langle\psi|\psi_\sigma\rangle|\psi\rangle\langle\psi_\sigma|\|_1 &= |\langle\psi_\sigma|\psi^*\rangle|\mathrm{Tr}\left[|\psi_\sigma\rangle\langle\psi_\sigma| + |\psi^*\rangle\langle\psi^*| - \langle\psi_\sigma|\psi^*\rangle|\psi_\sigma\rangle\langle\psi^*| - \langle\psi^*|\psi_\sigma\rangle|\psi^*\rangle\langle\psi_\sigma|\right] \\ &= |\langle\psi_\sigma|\psi^*\rangle|\mathrm{Tr}\left[(\sigma - |\psi^*\rangle\langle\psi^*|)(\sigma - |\psi^*\rangle\langle\psi^*|)^\dagger\right] \\ &= |\langle\psi_\sigma|\psi^*\rangle|\||\psi^*\rangle\langle\psi^*| - \sigma\|_1 \\ &= \delta(P_A, P_B)\sqrt{1 - \frac{\delta(P_A, P_B)^2}{4}}. \end{aligned} \tag{100}$$

Applying the reversed triangle inequality, we finally arrive at

$$\begin{aligned} r &\geq \left|\|\sigma - |\psi^*\rangle\langle\psi^*|\|_1 - \|2|\langle\psi_\sigma|\psi^*\rangle|^2\sigma - \langle\psi_\sigma|\psi^*\rangle|\psi_\sigma\rangle\langle\psi| - \langle\psi|\psi_\sigma\rangle|\psi\rangle\langle\psi_\sigma|\|_1\right| \\ &= \delta(P_A, P_B)\left(1 - \sqrt{1 - \frac{\delta(P_A, P_B)^2}{4}}\right). \end{aligned} \tag{101}$$

□

Supplementary Note 5. ROBUSTNESS WITH DEPOLARISED INPUT STATES

Corollary 1 (Depolarised single-qubit pure states). *Let $|\psi_\sigma\rangle, |\psi_\rho\rangle \in \mathbb{C}^2$ be single-qubit pure sates and let $\mathcal{E}_p^{\mathrm{dep}}$ be a depolarising channel with noise parameter $p \in (0, 1)$. Then, if $p_\mathrm{A} > 1/2$ and $p_\mathrm{B} = 1 - p_\mathrm{A}$, the robustness condition (10) for $\mathcal{E}_p^{\mathrm{dep}}(\sigma)$ and $\mathcal{E}_p^{\mathrm{dep}}(\rho)$ is equivalent to*

$$\frac{1}{2}\||\psi_\sigma\rangle\langle\psi_\sigma| - |\psi_\rho\rangle\langle\psi_\rho|\|_1 < \begin{cases} \sqrt{\frac{1}{2} - \sqrt{\frac{2p_\mathrm{A}(1-p_\mathrm{A}) - p(1-\frac{1}{2}p)}{2(1-p)^2}}} & \text{if } \frac{1}{2} < p_\mathrm{A} \leq \frac{4-6p+3p^2}{4-4p+2p^2}, \\ \sqrt{\frac{p\cdot(2-p)\cdot(1-2p_\mathrm{A})^2}{8(1-p)^2\cdot(1-p_\mathrm{A})}} & \text{if } \frac{4-6p+3p^2}{4-4p+2p^2} < p_\mathrm{A} \leq \frac{4-3p}{4-2p}, \\ 1 & \text{if} p_\mathrm{A} > \frac{4-3p}{4-2p}. \end{cases} \tag{102}$$

Proof. In order to prove the corollary, we proceed in a manner analogous to the proof of Lemma 1. Specifically, we show that the condition on the trace distance in eq. (102) is equivalent to the SDP robustness condition (10) from Theorem 1 expressed in terms of type-II error probabilities. Let $\sigma = |\psi_\sigma\rangle\langle\psi_\sigma|$, $\rho = |\psi_\rho\rangle\langle\psi_\rho|$ and recall that the type-I and type-II error probabilities are given by

$$\alpha(M;\, \mathcal{E}_p^{\mathrm{dep}}(\sigma)) = \mathrm{Tr}\left[M\mathcal{E}_p^{\mathrm{dep}}(\sigma)\right],\ \ \beta(M; \mathcal{E}_p^{\mathrm{dep}}(\rho)) = \mathrm{Tr}\left[(\mathbb{1} - M)\mathcal{E}_p^{\mathrm{dep}}(\rho)\right] \tag{103}$$

with $0 \leq M \leq \mathbb{1}_d$. Let $\sigma' := \mathcal{E}_p^{\mathrm{dep}}(\sigma)$ and $\rho' := \mathcal{E}_p^{\mathrm{dep}}(\rho)$ and recall that a Helstrom operator for testing the null σ' against the alternative ρ' with type-I error probability α_0 takes the form (60)

$$M_{\tau(\alpha_0)} := P_{\tau(\alpha_0),+} + q_0 P_{\tau(\alpha_0),0},\ \ q_0 := \begin{cases} \frac{\alpha_0 - \alpha(P_{\tau(\alpha_0),+})}{\alpha(P_{\tau(\alpha_0),0})} & \text{if } \alpha\left(P_{\tau(\alpha_0),0}\right) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{104}$$

where $\tau(\alpha_0) := \inf\{t \geq 0\colon\ \alpha(P_{t,+}) \leq \alpha_0\}$. Let $M_\mathrm{A}^\star := M_{\tau(1-p_\mathrm{A})}$ and $M_\mathrm{B}^\star := M_{\tau(p_\mathrm{B})}$ and note that by assumption $p_\mathrm{B} = 1 - p_\mathrm{A}$ and hence $M_\mathrm{A}^\star = M_\mathrm{B}^\star$. The SDP robustness condition then simplifies to $\beta_{1-p_\mathrm{A}}^*(\sigma', \rho') > 1/2$. We now proceed as follows. We first compute the spectral decomposition of the operator $\rho' - t\sigma'$ as a function of t and relate it to the fidelity between σ and ρ. With this, we derive an expression for $\alpha(P_{t,+})$ and subsequently compute $\tau(\alpha_0)$.

This yields an expression for the Helstrom operator with type-I error probability $1 - p_{\mathrm{A}}$ which can then be used to solve inequality (10) for the fidelity. We thus start by solving the eigenvalue problem

$$(\rho' - t\sigma')|\mu\rangle = \mu|\mu\rangle \tag{105}$$

which can be rewritten as

$$\left((1-p)\cdot(\rho - t\sigma) + \frac{p(1-t)}{2}\mathbb{1}_2\right)|\eta\rangle. \tag{106}$$

We notice that the operators $\rho' - t\sigma'$ and $\rho - t\sigma$ share the same set of eigenvectors. Furthermore, if η is an eigenvalue of $\rho - t\sigma$ with eigenvector $|\eta\rangle$, then the corresponding eigenvalue μ of $\rho' - t\sigma'$ is given by

$$\mu = (1-p)\eta + \frac{p\cdot(1-t)}{2}. \tag{107}$$

From the proof of Lemma 1, we know that the eigenvalues of $\rho - t\cdot\sigma$ are given by

$$\begin{gathered}\eta_0 = \frac{1}{2}(1-t) + R > 0, \;\; \eta_1 = \frac{1}{2}(1-t) - R \leq 0 \\ R = \sqrt{\frac{1}{4}(1-t)^2 + t(1-|\gamma|^2)}, \;\; \gamma = \langle\psi_\rho|\psi_\sigma\rangle\end{gathered} \tag{108}$$

with eigenvectors

$$\begin{gathered}|\eta_0\rangle = -\gamma A_0|\psi_\rho\rangle + (1-\eta_0)A_0|\psi_\sigma\rangle, \;\; |\eta_2\rangle = -\gamma A_2|\psi_\rho\rangle + (1-\eta_2)A_2|\psi_\sigma\rangle \\ |A_k|^{-2} = 2R\left|\eta_k - 1 + |\gamma|^2\right|.\end{gathered} \tag{109}$$

With this, we can compute the eigenvalues μ_k and eigenprojections P_k of $\rho' - t\sigma'$ as

$$\begin{gathered}\mu_0 = (1-p)\eta_0 + p\cdot\frac{1-t}{2}, \;\; \mu_1 = (1-p)\eta_1 + p\cdot\frac{1-t}{2}, \\ P_0 = |\eta_0\rangle\langle\eta_0|, \;\; P_1 = |\eta_1\rangle\langle\eta_1|.\end{gathered} \tag{110}$$

Since $\eta_0 > 0 \geq \eta_1$ for any $t \geq 0$, we have $\mu_0 \geq \mu_1$ and furthermore, the eigenvalues are monotonically decreasing functions of t for $|\gamma|^2 < 1$. To see this, consider

$$\frac{dR}{dt} = \frac{1 + t - 2|\gamma|^2}{2\sqrt{(1+t)^2 - 4t|\gamma|^2}} \tag{111}$$

and thus for $\forall\, t \geq 0$ and $|\gamma|^2 < 1$

$$\frac{d\mu_0}{dt} = \frac{dR}{dt} - \frac{1}{2} < 0 \;\text{ and }\; \frac{d\mu_1}{dt} = -\frac{dR}{dt} - \frac{1}{2} < 0. \tag{112}$$

Hence, since both eigenvalues are strictly positive at $t = 0$, there exists exactly one ξ_k such that μ_k vanishes at ξ_k, $k = 0, 1$. Algebra shows that these zeroes are given by

$$\begin{aligned}\xi_0 &= 1 + \frac{2(1-|\gamma|^2)(1-p)^2}{p(2-p)}\left(1 + \sqrt{1 + \frac{p(2-p)}{(1-|\gamma|^2)(1-p)^2}}\right) > 1, \\ \xi_1 &= 1 + \frac{2(1-|\gamma|^2)(1-p)^2}{p(2-p)}\left(1 - \sqrt{1 + \frac{p(2-p)}{(1-|\gamma|^2)(1-p)^2}}\right) < 1.\end{aligned} \tag{113}$$

We define the functions

$$g_0(t) := \langle \eta_0 | \sigma' | \eta_0 \rangle = \frac{1}{2} \left(1 + \frac{(1-p)(2\,|\gamma|^2 - 1 - t)}{\sqrt{(1+t)^2 - 4t\,|\gamma|^2}} \right), \tag{114}$$

$$g_1(t) := \langle \eta_1 | \sigma' | \eta_1 \rangle = \frac{1}{2} \left(1 - \frac{(1-p)(2\,|\gamma|^2 - 1 - t)}{\sqrt{(1+t)^2 - 4t\,|\gamma|^2}} \right), \tag{115}$$

$$f_0(t) := \langle \eta_0 | \rho' | \eta_0 \rangle = \frac{1}{2} \left(1 + \frac{(1-p)(1 + t \cdot (1 - 2\,|\gamma|^2)}{\sqrt{(1+t)^2 - 4t\,|\gamma|^2}} \right), \tag{116}$$

$$f_1(t) := \langle \eta_0 | \rho' | \eta_0 \rangle = \frac{1}{2} \left(1 - \frac{(1-p)(1 + t \cdot (1 - 2\,|\gamma|^2)}{\sqrt{(1+t)^2 - 4t\,|\gamma|^2}} \right). \tag{117}$$

With this, we now compute $t \mapsto \alpha(P_{t,+})$ as

$$\alpha(P_{t,+}) = \mathrm{Tr}\left[\sigma' P_{t,+}\right] = \begin{cases} 1 & 0 \le t < \xi_1 \\ g_0(t) & \xi_1 \le t < \xi_0 \\ 0 & \xi_0 \le t \end{cases} \tag{118}$$

For $\alpha_0 \in [0,\, 1]$, we compute $\tau(\alpha_0) := \inf\{t \ge 0\colon\ \alpha(P_{t,+}) \le \alpha_0\}$ as

$$\tau(\alpha_0) = \begin{cases} \xi_0 & 0 \le \alpha_0 \le g_0(\xi_0) \\ g_0^{-1}(\alpha_0) & g_0(\xi_0) < \alpha_0 < g_0(\xi_1) \\ \xi_1 & g_0(\xi_1) \le \alpha_0 < 1 \\ 0 & \alpha_0 = 1 \end{cases} \tag{119}$$

where

$$g_0^{-1}(\alpha_0) = 2\,|\gamma|^2 - 1 + 2(1 - 2\alpha_0)\sqrt{\frac{|\gamma|^2\,(1 - |\gamma|^2)}{p(2-p) - 4\alpha_0(1-\alpha_0)}}. \tag{120}$$

To solve condition (10) we now have to distinguish different cases, depending on which interval $1 - p_{\mathrm{A}}$ falls into. Firstly, if $1 - p_{\mathrm{A}} = 1$, then $\tau(1 - p_{\mathrm{A}}) = 0$ and thus $\beta(M_{\mathrm{A}}^\star) = 0$ in which case the condition can not be satisfied. If $1 - p_{\mathrm{A}} \in [g_0(\xi_1),\, 1)$, then we have $\tau(1 - p_{\mathrm{A}}) = \xi_1$. In this case, it holds that $\mu_0 > 0$ and $\mu_1 = 0$ and the Helstrom operator is given by

$$M_{\mathrm{A}}^\star = |\eta_0\rangle\langle\eta_0| + \frac{1 - p_{\mathrm{A}} - g_0(\xi_1)}{g_1(\xi_1)} |\eta_1\rangle\langle\eta_1| \tag{121}$$

and the robustness condition reads

$$\beta(M_{\mathrm{A}}^\star) = 1 - f_0(\xi_1) - \frac{1 - p_{\mathrm{A}} - g_0(\xi_1)}{g_1(\xi_1)} f_1(\xi_1) > \frac{1}{2} \tag{122}$$

which cannot to be satisfied simultaneously with $1 - p_{\mathrm{A}} \in [g_0(\xi_1),\, 1)$. If, on the other hand $1 - p_{\mathrm{A}} \in (g_0(\xi_0),\, g_0(\xi_1))$, then $\tau(1 - p_{\mathrm{A}}) = g_0^{-1}(1 - p_{\mathrm{A}})$ and $\mu_0 > 0 > \mu_1$. The Helstrom operator then is given by $M_{\mathrm{A}}^\star = |\eta_0\rangle\langle\eta_0|$ which leads to the robustness condition

$$1 - f_0(g_0^{-1}(1 - p_{\mathrm{A}})) > \frac{1}{2} \tag{123}$$

which, together with $1 - p_{\mathrm{A}} \in (g_0(\xi_0),\, g_0(\xi_1))$, is equivalent to

$$|\gamma|^2 > \frac{1}{2} \left(1 + \sqrt{\frac{4p_{\mathrm{A}}(1 - p_{\mathrm{A}}) - p(2-p)}{(1-p)^2}} \right), \quad \frac{1}{2} \le p_{\mathrm{A}} \le \frac{4 - 6p + 3p^2}{4 - 4p + 2p^2}. \tag{124}$$

In the last case where $1-p_{\mathrm{A}} \leq g_0(\xi_0)$, we have $\tau(1-p_{\mathrm{A}}) = \xi_0$ and thus $\mu_0 = 0 > \mu_1$. The Helstrom operator is then given by $\frac{1-p_{\mathrm{A}}}{g_0(\xi_0)}|\eta_0\rangle\langle\eta_0|$, leading to the robustness condition

$$1 - \frac{1-p_{\mathrm{A}}}{g_0(\xi_0)} f_0(\xi_0) > \frac{1}{2}. \tag{125}$$

Together with $1-p_{\mathrm{A}} \leq g_0(\xi_0)$ this is equivalent to

$$|\gamma|^2 > \begin{cases} \frac{4p_{\mathrm{A}}(1-p_{\mathrm{A}})-p(2-p)}{(1-p)^2(4(1-p_{\mathrm{A}})-p(2-p))}, & \text{if } \frac{1+(1-p)^2}{2} < p_{\mathrm{A}} \leq \frac{4-6p+3p^2}{4-4p+2p^2} \\ \frac{(4-3p-2p_{\mathrm{A}}(2-p))(2-p(3-2p_{\mathrm{A}}))}{8(1-p)^2(1-p_{\mathrm{A}})}, & \text{if } \frac{4-6p+3p^2}{4-4p+2p^2} < p_{\mathrm{A}} \leq \frac{4-3p}{4-2p}, \\ 0, & \text{if } p_{\mathrm{A}} > \frac{4-3p}{4-2p}. \end{cases} \tag{126}$$

Finally, combining together conditions (124) and (126) leads to

$$|\gamma|^2 > \begin{cases} \frac{1}{2}\left(1 + \sqrt{\frac{4p_{\mathrm{A}}(1-p_{\mathrm{A}})-p(2-p)}{(1-p)^2}}\right), & \text{if } \frac{1}{2} < p_{\mathrm{A}} \leq \frac{4-6p+3p^2}{4-4p+2p^2} \\ \frac{(4-3p-2p_{\mathrm{A}}(2-p))(2-p(3-2p_{\mathrm{A}}))}{8(1-p)^2(1-p_{\mathrm{A}})}, & \text{if } \frac{4-6p+3p^2}{4-4p+2p^2} < p_{\mathrm{A}} \leq \frac{4-3p}{4-2p}, \\ 0, & \text{if } p_{\mathrm{A}} > \frac{4-3p}{4-2p}. \end{cases} \tag{127}$$

Since by assumption ρ and σ are pure states the proof is completed by noting that we have

$$T(\rho,\, \sigma) = \sqrt{1 - F(\rho,\, \sigma)} \tag{128}$$

by the Fuchs-van de Graaf inequality. $\square$

Supplementary Note 6. ALGORITHMS

Here, we provide pseudocode for the algorithm presented in Section II E.

Protocol 1 Robustness Certification(σ, N, α, $\mathcal{A}$)

Require: Quantum state $\sigma \in \mathcal{S}(\mathcal{H})$, number of measurement shots N, error tolerance α, a quantum classifier $\mathcal{A} = (\mathcal{E},\, \{\Pi_k\}_{k\in\mathcal{C}})$.
Ensure: Predicted class k_A, prediction score p_{A} and robust radius r_F according to Eq. (17) in terms of fidelity.

Set counter $\mathbf{n}_k \leftarrow 0$ for every $k \in \mathcal{C}$.
for $k = 1, \ldots N$ **do**
  Apply quantum circuit $\mathcal{E}$ to initial state σ.
  Perform $|\mathcal{C}|$-outcome measurement $\{\Pi_k\}_{k\in\mathcal{C}}$ on the evolved state $\mathcal{E}(\sigma)$.
  Record measurement outcome k by setting $\mathbf{n}_k \leftarrow \mathbf{n}_k + 1$.
end for
Calculate empirical probability distribution $\hat{\mathbf{y}}_k^{(N)} \leftarrow \mathbf{n}_k N^{-1}$.
Extract the most likely class $k_A \leftarrow \arg\max_k \hat{\mathbf{y}}_k^{(N)}$.
Set $p_{\mathrm{A}} \leftarrow \hat{\mathbf{y}}_{k_A}^{(N)}(\sigma) - \sqrt{\frac{-\log(\alpha)}{2N}}$.
if $p_{\mathrm{A}} > 1/2$ **then**
  Calculate robust radius $r_F \leftarrow \frac{1}{2} + \sqrt{p_{\mathrm{A}}(1-p_{\mathrm{A}})}$.
  return prediction k_A, class score p_{A}, robust radius r_F.
else
  return ABSTAIN
end if

SUPPLEMENTARY REFERENCES

[1] Bhatia, R. *Matrix analysis*, vol. 169 (Springer, 2013).
[2] Richtmeyer, R. D. *Principles of Advanced Mathematical Physics*, vol. 1 (Springer, 1978).