PHYSICAL REVIEW
LETTERS

Statistical Distance and the Geometry of Quantum States

Samuel L. Braunstein* and Carlton M. Caves
*Center for Advanced Studies, Department of Physics and Astronomy, University of New Mexico,
Albuquerque, New Mexico 87131-1156*

By finding measurements that optimally resolve neighboring quantum states, we use statistical distinguishability to define a natural Riemannian metric on the space of quantum-mechanical density operators and to formulate uncertainty principles that are more general and more stringent than standard uncertainty principles.

One task of precision quantum measurements is to detect a weak signal that produces a small change in the state of some quantum system. Given an initial quantum state, the size of the signal parametrizes a path through the space of quantum states. Detecting a weak signal is thus equivalent to distinguishing neighboring quantum states along the path. We pursue this point of view by using the theory of parameter estimation to formulate the problem of distinguishing neighboring states. We find measurements that optimally resolve neighboring states, and we characterize their degree of distinguishability in terms of a Riemannian metric, increasing distance corresponding to more reliable distinguishability. These considerations lead directly to uncertainty principles that are more general and more stringent than standard uncertainty principles.

We begin by reviewing Wootters's derivation [1] of a distinguishability metric for probability distributions. After drawing N samples from a probability distribution, one can estimate the probabilities p_j as the observed frequencies f_j. The probability for the frequencies is given by a multinomial distribution, which for large N is proportional to a Gaussian $\exp[-(N/2)(f_j - p_j)^2/p_j]$. A nearby distribution $\tilde{p}_j$ can be reliably distinguished from p_j if the Gaussian $\exp[-(N/2)(\tilde{p}_j - p_j)^2/p_j]$ is small. Thus the quadratic form $(\tilde{p}_j - p_j)^2/p_j$ provides a natural (Riemannian) *distinguishability metric* on the space of probability distributions (PD),

$$ds^2_{\mathrm{PD}} \equiv \sum_j \frac{dp_j^2}{p_j} = \sum_j p_j (d\ln p_j)^2 = 4\sum_j dr_j^2 , \quad (1)$$

where $p_j = r_j^2$. Using this argument, Wootters was led to the distance s_{PD}, which he called *statistical distance*.

Wootters generalized statistical distance to quantum-mechanical pure states as follows [1]. Consider neighboring pure states, expanded in an orthonormal basis $|j\rangle$:

$$|\psi\rangle = \sum_j \sqrt{p_j} e^{i\varphi_j} |j\rangle , \quad (2)$$

$$|\tilde{\psi}\rangle = |\psi\rangle + |d\psi\rangle = \sum_j \sqrt{p_j + dp_j}\, e^{i(\varphi_j + d\varphi_j)} |j\rangle . \quad (3)$$

Normalization implies that $\mathrm{Re}(\langle\psi|d\psi\rangle) = -\frac{1}{2}\langle d\psi|d\psi\rangle$. Measurements described by the one-dimensional projectors $|j\rangle\langle j|$ can distinguish $|\psi\rangle$ and $|\tilde{\psi}\rangle$ according to the *classical* metric (1). The *quantum* distinguishability metric should be defined by measurements that resolve the two states optimally—i.e., that maximize Eq. (1).

The maximum is given by the Hilbert-space angle $\cos^{-1}(|\langle\tilde{\psi}|\psi\rangle|)$, which clearly captures a notion of state distinguishability. The corresponding line element,

$$\frac{1}{4}ds^2_{\mathrm{PS}} \equiv [\cos^{-1}(|\langle\tilde{\psi}|\psi\rangle|)]^2 = 1 - |\langle\tilde{\psi}|\psi\rangle|^2 = \langle d\psi_\perp|d\psi_\perp\rangle$$

$$= \frac{1}{4}\sum_j \frac{dp_j^2}{p_j} + \left[\sum_j p_j d\varphi_j^2 - \left(\sum_j p_j d\varphi_j\right)^2\right] , \quad (4)$$

called the Fubini-Study metric [2], is the natural metric on the manifold of Hilbert-space rays. Here $|d\psi_\perp\rangle \equiv |d\psi\rangle - |\psi\rangle\langle\psi|d\psi\rangle$ is the projection of $|d\psi\rangle$ orthogonal to $|\psi\rangle$. The term in large square brackets, the variance of the phase changes, is non-negative; an appropriate choice

of basis makes it zero [3]. Thus ds_{PS}^2 is the maximum value of ds_{PD}^2, which means that ds_{PS} is the statistical distance between neighboring pure states (PS).

We generalize the notion of statistical distance to impure quantum states and thus obtain a natural Riemannian geometry on the space of density operators [see Eqs. (23) and (28)], where no *natural* inner product guides the generalization. Our derivation, like Wootters's, proceeds in two steps, one classical and one quantum mechanical, but it sharpens the formulation of statistical distance by highlighting distinct classical and quantum optimization problems. For the first step, to obtain the classical distinguishability metric, we use an approach based on the theory of parameter estimation [4]. This approach, which is independent of Wootters's work, maps the problem of state distinguishability onto that of precision determination of a parameter. For the second step, to obtain the quantum distinguishability metric, we optimize over *all* quantum measurements, not just measurements described by one-dimensional orthogonal projectors.

Consider now a curve $\hat{\rho}(X)$ on the space of density operators. The problem of distinguishing $\hat{\rho}(X)$ from neighboring density operators along the curve is equivalent to the problem of determining the value of the parameter X. The determination is made from the results of measurements. To be general, we must allow arbitrary *generalized* quantum measurements [5,6], which include all measurements permitted by the rules of quantum mechanics.

A generalized measurement is described by a set of non-negative, Hermitian operators $\hat{E}(\xi)$, which are complete in the sense that

$$\int d\xi\, \hat{E}(\xi) = \hat{1} = (\text{unit operator}) .\qquad (5)$$

The quantity ξ labels the "results" of the measurement; although written here as a single continuous real variable, it could be discrete or multivariate. The probability density for result ξ, given the parameter X, is

$$p(\xi|X) = \text{tr}\big(\hat{E}(\xi)\hat{\rho}(X)\big) .\qquad (6)$$

Consider now N such measurements, with results $\xi_1,\ldots,\xi_N$. One estimates the parameter X via a function $X_{\text{est}} = X_{\text{est}}(\xi_1,\ldots,\xi_N)$. A sensible definition of statistical distance is to measure the parameter increment dX in units of the statistical deviation of the estimator away from the parameter. The appropriate measure of deviation is

$$\frac{X_{\text{est}}}{|d\langle X_{\text{est}}\rangle_X/dX|} - X \equiv \delta X .\qquad (7)$$

The derivative $d\langle X_{\text{est}}\rangle_X/dX$ removes the local difference in the "units" of the estimator and the parameter. The subscript X on expectation values reminds one that they depend on the parameter. The appropriate unit of sta-

tistical deviation is $\min\big[\sqrt{N}\langle(\delta X)^2\rangle_X^{1/2}\big]$. The $\sqrt{N}$ removes the expected $1/\sqrt{N}$ improvement with the number of measurements; the minimum means that statistical distance is defined in terms of the most discriminating procedure for determining the parameter.

We are thus led to define the distinguishability metric by

$$ds^2 \equiv \frac{dX^2}{\min\big[N\langle(\delta X)^2\rangle_X\big]} .\qquad (8)$$

We take the minimum in the two steps mentioned above: first, optimization over estimators for a given quantum measurement to get the classical distinguishability metric and, second, optimization over all quantum measurements to get the quantum distinguishability metric.

The classical optimization relies on a lower bound, called the Cramér-Rao bound [7], on the variance of any estimator. The proof of the Cramér-Rao bound proceeds from the trivial identity

$$0 = \int d\xi_1 \cdots d\xi_N\, p(\xi_1|X)\cdots p(\xi_N|X)\,\Delta X_{\text{est}} .\qquad (9)$$

where $\Delta X_{\text{est}} \equiv X_{\text{est}}(\xi_1,\ldots,\xi_N) - \langle X_{\text{est}}\rangle_X$. Taking the derivative of this identity with respect to X, we obtain

$$\int d\xi_1 \cdots d\xi_N\, p(\xi_1|X)\cdots p(\xi_N|X)$$
$$\times \left(\sum_{n=1}^{N} \frac{\partial \ln p(\xi_n|X)}{\partial X}\right)\Delta X_{\text{est}} = \frac{d\langle X_{\text{est}}\rangle_X}{dX} .\qquad (10)$$

Applying the Schwarz inequality to Eq. (10) yields the *Cramér-Rao bound*

$$NF(X)\langle(\Delta X_{\text{est}})^2\rangle_X \geq \left(\frac{d\langle X_{\text{est}}\rangle_X}{dX}\right)^2 ,\qquad (11)$$

where the *Fisher information* is defined by

$$F(X) \equiv \int d\xi\, p(\xi|X)\left(\frac{\partial \ln p(\xi|X)}{\partial X}\right)^2$$
$$= \int d\xi\, \frac{1}{p(\xi|X)}\left(\frac{\partial p(\xi|X)}{\partial X}\right)^2 .\qquad (12)$$

Converted to the form needed in the definition (8), the Cramér-Rao bound becomes

$$N\langle(\delta X)^2\rangle_X \geq \frac{1}{F(X)} + N\langle\delta X\rangle_X^2 \geq \frac{1}{F(X)} .\qquad (13)$$

A nonzero value of $\langle\delta X\rangle_X$ means that the units-corrected estimator has a systematic bias away from the parameter; $\langle\delta X\rangle_X$ is zero when the estimator is unbiased, i.e., when $\langle X_{\text{est}}\rangle_X = X$ locally.

The Cramér-Rao bound only places a lower bound on the minimum that appears in Eq. (8). Fisher's theorem [8,9], however, says that asymptotically for large N, maximum-likelihood estimation is unbiased and achieves the Cramér-Rao bound. Thus, for a given probabil-

ity distribution $p(\xi|X)$, we arrive at the classical distinguishability metric $ds_{\text{PD}}^2 = F(X)dX^2$, which, given the forms (12) of F, is the Wootters metric (1) for continuous, instead of discrete alternatives.

The second step, to optimize over quantum measurements, is now seen to be the problem of maximizing the Fisher information over all quantum measurements, i.e., symbolically

$$ds_{\text{DO}}^2 = dX^2 \max_{\{\hat{E}(\xi)\}} F(X) . \tag{14}$$

The subscript DO reminds one that this is a metric on density operators.

The expression for $F(X)$ involves dividing by $p(\xi|X)$, so one might expect the quantum distinguishability metric to involve "division" by $\hat{\rho}$. The appropriate sense of this division comes from defining a superoperator

$$\mathcal{R}_{\hat{\rho}}(\hat{O}) \equiv \tfrac{1}{2}(\hat{\rho}\hat{O} + \hat{O}\hat{\rho}) = \sum_{j,k} \tfrac{1}{2}(p_j + p_k)O_{jk}|j\rangle\langle k| . \tag{15}$$

The second form is written in the orthonormal basis where $\hat{\rho} = \sum_j p_j|j\rangle\langle j|$ is diagonal. In the interior of the space of density operators—i.e., away from the boundary, where one or more of the eigenvalues p_j vanishes—$\mathcal{R}_{\hat{\rho}}$ has a well defined inverse $\mathcal{R}_{\hat{\rho}}^{-1}$, with matrix elements $[\mathcal{R}_{\hat{\rho}}^{-1}(\hat{O})]_{jk} = 2O_{jk}/(p_j + p_k)$ in the basis that diagonalizes $\hat{\rho}$. The only property of $\mathcal{R}_{\hat{\rho}}^{-1}$ we need is that for Hermitian $\hat{A}$ and $\hat{B}$,

$$\text{tr}(\hat{A}\hat{B}) = \text{Re}[\text{tr}(\hat{\rho}\hat{A}\mathcal{R}_{\hat{\rho}}^{-1}(\hat{B}))] . \tag{16}$$

To proceed, we put the quantum probability distribution (6) into the Fisher information (12) to obtain

$$F(X) = \int d\xi \frac{[\text{tr}(\hat{E}(\xi)\hat{\rho}'(X))]^2}{\text{tr}(\hat{E}(\xi)\hat{\rho}(X))} , \tag{17}$$

where $\hat{\rho}'(X) \equiv d\hat{\rho}/dX$. In the integrand we now substitute, using property (16) with $\hat{A} = \hat{E}$ and $\hat{B} = \hat{\rho}'$. Since this substitution introduces $\mathcal{R}_{\hat{\rho}}^{-1}$, it is instructive to enquire into its validity on the boundary.

The enquiry begins by writing $\hat{\rho}$ and $\hat{\rho} + dX\hat{\rho}'$ in their orthonormal bases,

$$\hat{\rho} = \sum_j p_j|j\rangle\langle j| , \tag{18}$$

$$\hat{\rho} + dX\hat{\rho}' = \sum_j (p_j + dp_j)|j'\rangle\langle j'|$$

$$= \sum_j dp_j|j\rangle\langle j| + e^{idX\hat{h}}\hat{\rho}e^{-idX\hat{h}} , \tag{19}$$

where the Hermitian operator $\hat{h}$ generates the infinitesimal unitary basis transformation:

$$|j'\rangle = e^{idX\hat{h}}|j\rangle = \sum_k (\delta_{kj} + idXh_{kj})|k\rangle . \tag{20}$$

The analog of the coordinate singularity in the Woot-

ters metric (1) at the boundary can be removed by using coordinates r_j, where $p_j = r_j^2$, which essentially remove the boundary. One now shows that

$$dX\hat{\rho}' = \sum_j dp_j|j\rangle\langle j| + idX\sum_{j,k}(p_j - p_k)h_{kj}|k\rangle\langle j| , \tag{21}$$

from which it follows that

$$dX\text{tr}(\hat{A}\hat{\rho}') = 2\sum_j r_j A_{jj}dr_j + idX\sum_{j,k}(p_j - p_k)A_{jk}h_{kj}$$

$$= dX\text{Re}[\text{tr}(\hat{\rho}\hat{A}\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}'))] , \tag{22}$$

provided that the singular matrix elements of $\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')$ are assigned any finite values consistent with Hermiticity. Choosing them to vanish conveniently extends $\mathcal{R}_{\hat{\rho}}^{-1}$ to the boundary

$$\mathcal{R}_{\hat{\rho}}^{-1}(\hat{O}) \equiv \sum_{\{j,k|p_j+p_k\neq0\}} \frac{2}{p_j + p_k}O_{jk}|j\rangle\langle k| . \tag{23}$$

We now manipulate the Fisher information (17) to obtain an upper bound

$$F = \int d\xi \frac{\left(\text{Re}[\text{tr}(\hat{\rho}\hat{E}(\xi)\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}'))]\right)^2}{\text{tr}(\hat{E}(\xi)\hat{\rho})}$$

$$\leq \int d\xi \frac{|\text{tr}(\hat{\rho}\hat{E}(\xi)\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}'))|^2}{\text{tr}(\hat{E}(\xi)\hat{\rho})} \tag{I}$$

$$= \int d\xi \left| \text{tr}\left(\frac{\hat{\rho}^{1/2}\hat{E}^{1/2}(\xi)}{\sqrt{\text{tr}(\hat{E}(\xi)\hat{\rho})}} \hat{E}^{1/2}(\xi)\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')\hat{\rho}^{1/2} \right) \right|^2$$

$$\leq \int d\xi \, \text{tr}(\hat{E}(\xi)\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')\hat{\rho}\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')) \tag{II}$$

$$= \text{tr}(\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')\hat{\rho}\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')) . \tag{24}$$

Step (II) relies on the Schwarz inequality $|\text{tr}(\hat{O}^\dagger\hat{P})|^2 \leq \text{tr}(\hat{O}^\dagger\hat{O})\text{tr}(\hat{P}^\dagger\hat{P})$, and the final step follows from the completeness property (5).

The necessary and sufficient conditions for equality in Eq. (24) are, from step (I),

$$\text{Im}[\text{tr}(\hat{\rho}\hat{E}(\xi)\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}'))] = 0 \quad \text{for all } \xi , \tag{25}$$

and, from use of the Schwarz inequality in step (II),

$$\hat{E}^{1/2}(\xi)\hat{\rho}^{1/2} = \lambda_\xi \hat{E}^{1/2}(\xi)\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')\hat{\rho}^{1/2} \quad \text{for all } \xi , \tag{26}$$

where $\lambda_\xi = \text{tr}(\hat{E}(\xi)\hat{\rho})/\text{tr}(\hat{\rho}\hat{E}(\xi)\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}'))$ is a constant that depends only on ξ. Notice that condition (25) is equivalent to the requirement that λ_ξ be real.

In the interior of density-operator space, conditions (25) and (26) are equivalent to

$$\hat{E}^{1/2}(\xi)[\hat{1} - \lambda_\xi\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')] = 0 \quad \text{for all } \xi , \tag{27}$$

with λ_ξ real. On the boundary, condition (27) is sufficient, but not necessary. Condition (27) means that $\hat{E}^{1/2}(\xi)$ and, hence, $\hat{E}(\xi)$ act within a single degenerate subspace of $\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')$, with λ_ξ being the inverse of the eigenvalue of $\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')$ within that subspace. This condition can always be met by choosing the operators $\hat{E}(\xi)$ to be one-dimensional projectors onto a complete set of orthonormal eigenstates of $\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')$.

The upper bound (24) thus being achievable, the distinguishability metric (14) on density operators becomes

$$ds_{\mathrm{DO}}^2 = \mathrm{tr}\big(\mathcal{R}_{\hat{\rho}}^{-1}(d\hat{\rho})\hat{\rho}\mathcal{R}_{\hat{\rho}}^{-1}(d\hat{\rho})\big) = \mathrm{tr}\big(d\hat{\rho}\mathcal{R}_{\hat{\rho}}^{-1}(d\hat{\rho})\big) , \quad (28)$$

where the second form follows from Eq. (22) with $\hat{A} = \mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')$. We stress that an unachievable upper bound cannot be used to define statistical distance.

On the pure-state boundary, where $\hat{\rho} = |\psi\rangle\langle\psi|$ and $dX\hat{\rho}' = d\hat{\rho} = |\psi\rangle\langle d\psi_\perp| + |d\psi_\perp\rangle\langle\psi| = \frac{1}{2}\mathcal{R}_{\hat{\rho}}^{-1}(d\hat{\rho})$, the density-operator metric (28) reduces to the Fubini-Study metric (4): $ds_{\mathrm{DO}}^2 = 2\mathrm{tr}\big((d\hat{\rho})^2\big) = 4\langle d\psi_\perp|d\psi_\perp\rangle = ds_{\mathrm{PS}}^2$. The conditions (25) and (26) for optimal measurements become

$$\mathrm{Im}[\langle\psi|\hat{E}(\xi)|d\psi_\perp\rangle] = 0 \quad \text{for all } \xi , \quad (29)$$

$$\hat{E}^{1/2}(\xi)(dX|\psi\rangle - 2\lambda_\xi|d\psi_\perp\rangle) = 0 \quad \text{for all } \xi . \quad (30)$$

These conditions mean that some linear combination of $|\psi\rangle$ and $|d\psi_\perp\rangle$, with *real* coefficients, is a zero-eigenvalue eigenstate of $\hat{E}^{1/2}(\xi)$ and, hence, of $\hat{E}(\xi)$. If the operators $\hat{E}(\xi)$ form a complete set of one-dimensional orthogonal projectors $|j\rangle\langle j|$, condition (29) implies condition (30), so the two conditions reduce to $\mathrm{Im}(\langle\psi|j\rangle\langle j|d\psi_\perp\rangle) = 0$ for all j, which can always be satisfied [3].

Both Helstrom [10] and Holevo [6] have derived the bound that comes from combining Eqs. (11) and (24), with $\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')$ called the "symmetric logarithmic derivative." Our procedure reaches the ultimate bound (24) through two uses of the Schwarz inequality, the first to get the classical Cramér-Rao bound (11) and the second to get the quantum bound (24). Helstrom and Holevo proceed directly to the quantum bound (24) through a single Schwarz inequality applied to a more complicated operator inner product. Their procedure obscures the separate classical and quantum optimization problems, thus making it difficult to investigate whether the bound is achievable, a question neither addresses.

Interestingly, the density-operator metric (28) has appeared in another context. Bures [11] defined a distance between density operators, which Uhlmann [12] interpreted as a generalization of transition probabilities to mixed states. Uhlmann found an explicit form for the Bures distance [13],

$$d_{\mathrm{Bures}}(\hat{\rho}_1, \hat{\rho}_2) = \sqrt{2}\big[1 - \mathrm{tr}\big((\hat{\rho}_1^{1/2}\hat{\rho}_2\hat{\rho}_1^{1/2})^{1/2}\big)\big]^{1/2} , \quad (31)$$

which for neighboring density operators reduces to [13]

$$[d_{\mathrm{Bures}}(\hat{\rho}, \hat{\rho} + d\hat{\rho})]^2 = \tfrac{1}{4}\mathrm{tr}\big(d\hat{\rho}\mathcal{R}_{\hat{\rho}}^{-1}(d\hat{\rho})\big) = \tfrac{1}{4}ds_{\mathrm{DO}}^2.$$

Combining Eqs. (13) and (14) yields an uncertainty principle for estimating the parameter X,

$$N\langle(\delta X)^2\rangle_X \frac{ds_{\mathrm{DO}}^2}{dX^2} \geq 1 . \quad (32)$$

where

$$\frac{ds_{\mathrm{DO}}^2}{dX^2} = \langle\big(\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')\big)^2\rangle_X = \mathrm{tr}\big(\hat{\rho}'\mathcal{R}_{\hat{\rho}}^{-1}(\hat{\rho}')\big)$$

$$= \sum_j \frac{(dp_j/dX)^2}{p_j} + 2\sum_{j,k} \frac{(p_j - p_k)^2}{p_j + p_k}|h_{jk}|^2 \quad (\mathrm{I}')$$

$$\leq \sum_j \frac{(dp_j/dX)^2}{p_j} + 4\langle(\Delta\hat{h})^2\rangle_X . \quad (\mathrm{II}')$$

$$(33)$$

Form (I') is written in the basis that diagonalizes $\hat{\rho}$; the infinitesimal probability changes dp_j and the Hermitian generator $\hat{h}$ are as in Eqs. (19) and (20). In form (I') $\hat{h}$ can be replaced by $\Delta\hat{h} = \hat{h} - \langle\hat{h}\rangle_X$; form (II'), involving the variance $\langle(\Delta\hat{h})^2\rangle_X$ of $\hat{h}$, follows immediately. Equality holds in step (II') if and only if $p_j p_k|\Delta h_{jk}|^2 = 0$ for all j and k; in particular, equality always holds if $\hat{\rho}$ is a pure state, but never holds if $\hat{\rho}$ is in the interior of density-operator space (except in the trivial case $\Delta\hat{h} = 0$).

Forms (I') and (II') interpolate smoothly between a "classical" uncertainty principle, when $\hat{h} = 0$, which limits distinguishability of probability distributions, and a "quantum" uncertainty principle, when $dp_j = 0$ for all j, which involves the generator $\hat{h}$. Even in the quantum case, the uncertainty principle is more general than standard uncertainty principles, because it allows many measurements and it is a Mandelstam-Tamm uncertainty principle [6,10] for a parameter X and the "conjugate" operator $\hat{h}$. At the same time, in the interior of density-operator space, form (I') is more stringent than standard uncertainty principles, which involve the variance of $\hat{h}$.

Current address: Department of Chemical Physics, Weizmann Institute of Science, 76 100 Rehovot, Israel.

[1] W. K. Wootters, Phys. Rev. D **23**, 357 (1981). Wootters's distance is half the distance we use here.

[2] G. W. Gibbons, J. Geom. Phys. **8**, 147 (1992).

[3] The condition is that $0 = p_j(d\varphi_j - \sum_k p_k d\varphi_k) = \mathrm{Im}(\langle\psi|j\rangle\langle j|d\psi_\perp\rangle)$ for all j. If a basis meets this condition, the basis vectors can always be rephased so that both $\langle j|\psi\rangle$ and $\langle j|d\psi_\perp\rangle$ are real. In particular, any basis that includes $|\psi\rangle$ or $|d\psi_\perp\rangle/\frac{1}{2}ds_{\mathrm{PS}}$ meets the condition.

[4] L. L. Campbell, Inform. Sci. **35**, 199 (1985).

[5] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).

[6] A. S. Holevo, *Probabilistic and Statistical Aspects of*

Quantum Theory (North-Holland, Amsterdam, 1982), especially Chaps. III.2 and VI.2.

[7] H. Cramér, *Mathematical Methods of Statistics* (Princeton University, Princeton, NJ, 1946), p. 500.

[8] R. A. Fisher, Proc. Camb. Soc. **22**, 700 (1925).

[9] S. L. Braunstein, J. Phys. A **25**, 3813 (1992); Phys. Rev. Lett. **69**, 3598 (1992).

[10] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976), Chap. VIII.4.

[11] D. J. C. Bures, Trans. Am. Math. Soc. **135**, 199 (1969).

[12] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).

[13] M. Hübner, Phys. Lett. A **163**, 239 (1992).