Quantum limit to subdiffraction incoherent optical imaging. II. A parametric-submodel approach

Mankei Tsang ●*

*Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583
and Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117551*

In a previous paper [M. Tsang, Phys. Rev. A **99**, 012305 (2019)], I proposed a quantum limit to the estimation of object moments in subdiffraction incoherent optical imaging. In this sequel, I prove the quantum limit rigorously by infinite-dimensional analysis. A key to the proof is the choice of an unfavorable parametric submodel to give a bound for the semiparametric problem. By generalizing the quantum limit for a larger class of moments, I also prove that the measurement method of spatial-mode demultiplexing with just one or two modes is able to achieve the quantum limit. For comparison, I derive a classical bound for direct imaging using the parametric-submodel approach, which suggests that direct imaging is substantially inferior.

I. INTRODUCTION

No problem is more essential in optics than the resolution limit of incoherent imaging [1]. Its importance to astronomy [2], fluorescence microscopy [3], and countless other imaging applications can hardly be overestimated. The fact that diffraction and photon shot noise play dual roles in limiting the ultimate resolution suggests that quantum information theory sets the right foundation for the problem [4,5]. Nevertheless, the mathematics of quantum information is daunting, its application to imaging more so, and progress has been mostly limited to toy examples, in which only a few parameters or hypotheses about the object are assumed to be unknown [4,5].

In recent years, the surprising results concerning two point sources in Ref. [6] have triggered renewed interest in the quantum-information perspective [5], as well as research efforts towards more general cases [7–23]. In terms of computing the quantum limits, some researchers have attacked the case of a few point sources and many parameters with numerical methods [17,18], but the harsh computational demands mean that alternative approaches are necessary to deal with more complex objects and high-dimensional parameter spaces. In this regard, Refs. [7–14] have been able to make progress by framing the problem as object-moment estimation, while assuming little about the object distribution. The prequel to this paper in particular proposes a quantum limit to moment estimation in the form of a quantum Cramér-Rao bound [12].

Two mathematical issues arise in the moment estimation problem: the infinite dimensionality of the parameter space, since an extended object may depend on infinitely many scalar parameters, and the infinite dimensionality of the quantum states, since an extended object may excite infinitely many spatial modes. The prequel ignored these issues and relied on finite-dimensional arguments. The main goal of this paper is to prove the quantum limit rigorously by infinite-dimensional analysis [24–35].

The theory of quantum semiparametric estimation, recently proposed in Ref. [36], is a key to the proof. Although the parameter space is infinite dimensional, the theory enables one to derive lower error bounds by considering parametric submodels, each of which depends on just a scalar parameter and is much easier to handle.

The second goal of this paper is to prove the quantum optimality of a measurement method called spatial-mode demultiplexing (SPADE) [5,6] for unbiased moment estimation. Previous works fixated on the estimation of simple moments [in the form of $\int x^\mu P(dx)$ for an object distribution P and a positive integer μ]. To estimate a simple moment without bias, measurements of infinitely many spatial modes are needed and it becomes difficult to even prove that an unbiased estimator exists [8–10,12]. In practice, of course, only a finite number of modes can be measured [37,38], so the existing results do not reflect well on the optimality of SPADE in practice. This paper generalizes the quantum limit for a larger class of moments so that, with just one or two modes, SPADE can still achieve the quantum limit for the generalized moments it is naturally measuring. There does not seem to be any compelling reason in practice to prefer a simple moment over a generalized moment pretty close to it, so there is arguably little loss of practical relevance and much to gain in the rigor by generalizing the moments.

The final goal of this paper is to give a bound for generalized moment estimation with direct imaging, thus proving the superiority of SPADE. While a similar result was proposed in Refs. [8–10], it relies on a special assumption about the point-spread function that is hard to check and has been verified only for a Gaussian point-spread function. I attempt to relax the assumption by appealing again to the parametric-submodel approach. Although the result is still not as general as one would like, it at least establishes conditions that are easier to check and paves the way for further generalizations.

_____

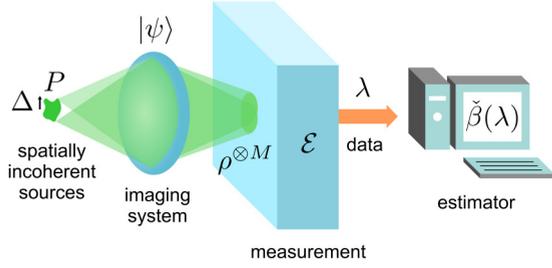*mankei@nus.edu.sg; https://blog.nus.edu.sg/mankei/

FIG. 1. Schematic of an imaging system. See the text for the definitions of the symbols.

This paper is organized as follows. Section II introduces a quantum model of incoherent optical imaging. Section III introduces the quantum semiparametric estimation theory. Section IV proposes a parametric submodel for the derivation of the quantum limit. Section V presents the quantum limit. Section VI proves that SPADE with one or two modes can still achieve the quantum limit. Section VII gives a bound for direct imaging. Section VIII discusses the conclusions. The Appendixes contain the more technical proofs and remarks.

II. MODEL

Let $\mathcal{H}_0$ be a one-dimensional (1D) Hilbert space for the vacuum, $\mathcal{H}$ a separable Hilbert space that models the spatial modes of light, τ_0 the vacuum state on $\mathcal{H}_0$, and τ a one-photon state on $\mathcal{H}$. Suppose that the state in each temporal mode can be modeled as

$$\rho = (1 - \epsilon)\tau_0 \oplus \epsilon\tau, \qquad (2.1)$$

where $\oplus$ denotes the direct sum and ϵ is the one-photon probability per temporal mode. With M temporal modes, the state is assumed to be the tensor power $\rho^{\otimes M}$ and the expected photon number in all modes is

$$N \equiv M\epsilon. \qquad (2.2)$$

The validity of this rare-photon model with $\epsilon \ll 1$ to describe thermal light at optical frequencies was studied extensively in Refs. [5,6,12]. Taking the limit $\epsilon \to 0$ while keeping N fixed leads to a Poisson model [39] that agrees with semiclassical optics [2], although it is not necessary to consider the Poisson limit in the following.

Assume 1D imaging for simplicity. Let $\mathbb{R}$ be the set of object-plane coordinates and Σ the Borel sigma algebra of $\mathbb{R}$ [31]. For an object that emits spatially incoherent light and is imaged with a diffraction-limited system, τ can be modeled as [6,12]

$$\tau(P) = \int e^{-i\hat{k}x} |\psi\rangle \langle\psi| e^{i\hat{k}x} P(dx), \qquad (2.3)$$

where $P : \Sigma \to [0, 1]$ is a probability measure that models the object distribution normalized by the total brightness, $|\psi\rangle \in \mathcal{H}$ with $\langle\psi|\psi\rangle = 1$ models the coherent point-spread function of the imaging system, and $\hat{k}$ is a self-adjoint operator on $\mathcal{H}$ for the optical spatial frequency. I call any self-adjoint operator an observable in the following. Figure 1 illustrates the imaging system. To work with the probability space $(\mathbb{R}, \Sigma, P)$, it will be useful to define an inner product, weighted by P, between two real functions $u, v : \mathbb{R} \to \mathbb{R}$ as

$$\langle u, v \rangle_P \equiv \int u(x)v(x)P(dx), \qquad (2.4)$$

the corresponding norm as

$$\|u\|_P \equiv \sqrt{\langle u, u \rangle_P}, \qquad (2.5)$$

and the resulting real Hilbert space as [31]

$$L_2(P) \equiv \{u : \|u\|_P < \infty\}. \qquad (2.6)$$

I assume $L_2(P)$ to be separable in the following.

The $\hat{k}$ and $|\psi\rangle$ lead to another probability space. Let the spectral representation of $\hat{k}$ be [29]

$$\hat{k} = \int k E(dk), \qquad (2.7)$$

where E is a projection-valued measure on $(\mathbb{R}, \Sigma)$. Define the spatial-frequency measure with respect to $\hat{k}$ and $|\psi\rangle$ as

$$Q(\cdot) \equiv \langle\psi| E(\cdot) |\psi\rangle. \qquad (2.8)$$

For example, if $\hat{k}$ is a continuous variable, then $E(dk) = |k\rangle \langle k| \, dk$, with $\langle k|k'\rangle = \delta(k - k')$ in the Dirac notation, and $Q(dk) = |\langle k|\psi\rangle|^2 dk$, where $\langle k|\psi\rangle$ is the optical transfer function of the imaging system.

In quantum information theory [40,41], it is often useful to find a purification of τ in a larger Hilbert space $\mathcal{H} \otimes \mathcal{H}'$ such that

$$\tau = \text{tr}'\Psi, \quad \Psi \equiv |\Psi\rangle \langle\Psi|, \quad |\Psi\rangle \in \mathcal{H} \otimes \mathcal{H}', \qquad (2.9)$$

where tr' denotes the partial trace over $\mathcal{H}'$. A natural choice is to take $\mathcal{H}' = L_2(P)$ using the representation

$$1 \in L_2(P) \leftrightarrow |\phi\rangle \in \mathcal{H}', \qquad (2.10)$$

$$u(x) \in L_2(P) \leftrightarrow u(\hat{x}) |\phi\rangle \in \mathcal{H}', \qquad (2.11)$$

where $\hat{x}$ is called the canonical multiplication operator, which is self-adjoint [see [33], Exercise 12.7(iii)], and $|\phi\rangle$ is called a cyclic vector, satisfying $\langle\phi|\phi\rangle = 1$ [29]. To adhere to physics terminology, I call $|\phi\rangle$ a purification of P. The purification of τ then becomes

$$|\Psi\rangle = e^{-i\hat{k}\otimes\hat{x}} |\psi\rangle \otimes |\phi\rangle. \qquad (2.12)$$

Let the spectral representation of $\hat{x}$ be

$$\hat{x} = \int x E'(dx), \qquad (2.13)$$

where E' is a projection-valued measure on $(\mathbb{R}, \Sigma)$. Then

$$P(\cdot) = \langle\phi| E'(\cdot) |\phi\rangle. \qquad (2.14)$$

For example, if P can be expressed in terms of a density $f(x)$ with respect to the Lebesgue measure, then $|\phi\rangle$ should satisfy, in the Dirac notation, $P(dx) = f(x)dx = \langle\phi| E'(dx) |\phi\rangle = |\langle x|\phi\rangle|^2 dx$. In other words, given an f, the purification should have a wave function that satisfies $|\langle x|\phi\rangle| = \sqrt{f(x)}$.

The τ can be purified in other ways by applying isometries on $\mathcal{H}'$ to $|\Psi\rangle$. Let $U : \mathcal{H}' \to \mathcal{H}''$ be an isometry that satisfies $U^\dagger U = I_{\mathcal{H}'}$, $I_{\mathcal{H}'}$ being the identity operator on $\mathcal{H}'$. Then an alternative purification is

$$|\Phi\rangle = I_{\mathcal{H}} \otimes U |\Psi\rangle, \quad \tau = \text{tr}''\Phi, \quad \Phi \equiv |\Phi\rangle \langle\Phi|. \qquad (2.15)$$

A fruitful choice made in Ref. [12] is as follows.

Lemma 1. Consider the state given by Eq. (2.3). Assume that the support of P [32], denoted by supp$P$, is infinite but bounded, viz.,

$$\#\mathrm{supp}P = \infty, \tag{2.16}$$

$$\Delta \equiv \sup_{x \in \mathrm{supp}P} |x| \in (0, \infty). \tag{2.17}$$

Assume also that $\hat{k}$ is bounded. Then a purification in $\mathcal{H} \otimes \mathcal{H}''$ is given by

$$|\Phi\rangle = \sum_{p=0}^{\infty} \sum_{n=0}^{p} \frac{(-i\hat{k})^p}{p!} L_{pn} |\psi\rangle \otimes |n\rangle, \tag{2.18}$$

where $\{|n\rangle : n \in \mathbb{N}_0\}$ is an orthonormal sequence in $\mathcal{H}''$ and L is the lower triangular matrix obtained by applying the Cholesky factorization algorithm [42] to the Hankel matrix

$$H_{pq} \equiv \langle x^p, x^q \rangle_P, \quad p, q \in \mathbb{N}_0. \tag{2.19}$$

The infinite series in Eq. (2.18) converges strongly (see [25], Definition 3.3.9).

The proof is deferred to Appendix A.

The assumptions about P ensure that H and L are defined. The bounded $\hat{k}$ is a technical assumption to ensure the convergence of results. Physically, the $\#\mathrm{supp}P = \infty$ assumption means that the object consists of infinitely many point sources; in other words, it is modeled as an extended object [2]. The bounded support simply means that the object has a finite size, while a bounded $\hat{k}$ simply means that the imaging system has a finite bandwidth.

III. SEMIPARAMETRIC ESTIMATION

Before proceeding further with the imaging problem, I review the quantum semiparametric theory proposed in Ref. [36], which is necessary to deal with the infinite-dimensional parameter space of the problem.

Let the statistical model of quantum states on a Hilbert space $\mathcal{H}$ be $\mathbf{G} \equiv \{\rho(g) : g \in \mathcal{G}\}$, where $\mathcal{G}$ is a possibly infinite-dimensional parameter space, and let the parameter of interest be a scalar $\beta : \mathcal{G} \to \mathbb{R}$. Let $\rho \in \mathbf{G}$ be the true state. Define an inner product, weighted by ρ, between two observables u and v as

$$\langle u, v \rangle_\rho \equiv \mathrm{tr}(u \circ v)\rho, \tag{3.1}$$

where $u \circ v \equiv (uv + vu)/2$ defines the Jordan product. The corresponding norm is

$$\|u\|_\rho \equiv \sqrt{\langle u, u \rangle_\rho}. \tag{3.2}$$

With respect to this inner product, define the real Hilbert space $L_2(\rho)$ as the completion of the set $\mathcal{B}(\mathcal{H})$ of bounded observables [34,35]. Within $L_2(\rho)$, define a subspace of zero-mean observables as

$$\mathcal{Z} \equiv \{u \in L_2(\rho) : \langle I_\mathcal{H}, u \rangle_\rho = 0\}. \tag{3.3}$$

Consider a 1D submodel, containing the true ρ, given by

$$\{\sigma(\theta) : \theta \in \mathcal{R} \subseteq \mathbb{R}, \sigma(\theta_0) = \rho\} \subseteq \mathbf{G}, \tag{3.4}$$

where $\mathcal{R}$ is an open interval containing θ_0. Let the overdot denote the derivative with respect to the parameter that is evaluated at the truth, such as

$$\dot{\sigma} \equiv \left. \frac{\partial \sigma(\theta)}{\partial \theta} \right|_{\theta=\theta_0}. \tag{3.5}$$

Assume that the submodel is regular, as defined by the conditions that $\sigma(\theta)$ is differentiable at the truth (such that $\dot{\sigma}$ is trace class [34]) and $|\mathrm{tr}u\dot{\sigma}| \leqslant C\|u\|_\rho$ for all $u \in \mathcal{B}(\mathcal{H})$ and some constant C. Then the bounded linear functional $\mathrm{tr}u\dot{\sigma}$ of u is continuous with respect to the norm $\|u\|_\rho$ (see [25], Theorem 1.5.7) and can be extended uniquely to be defined on $L_2(\rho)$ (see [25], Theorem 1.5.10). By the Riesz representation theorem (see [25], Theorem 3.7.7), there exists a unique $S^\sigma \in L_2(\rho)$, called the score, such that

$$\mathrm{tr}u\dot{\sigma} = \langle u, S^\sigma \rangle_\rho \, \forall u \in L_2(\rho). \tag{3.6}$$

This abstract definition of the score is due to Holevo [34,35]. To put it another way, suppose that an observable $\hat{S}^\sigma$, called a symmetric logarithmic derivative, is a solution to the Lyapunov equation

$$\dot{\sigma} = \hat{S}^\sigma \circ \rho \tag{3.7}$$

and $\|\hat{S}^\sigma\|_\rho < \infty$. Then $\mathrm{tr}u\dot{\sigma} = \langle u, \hat{S}^\sigma \rangle_\rho$ for all $u \in \mathcal{B}(\mathcal{H})$ by Eq. (2.8.88) in Ref. [34] and $\hat{S}^\sigma$ must be in the equivalence class of S^σ by the uniqueness of S^σ. In other words, a regular submodel is defined by having finite Helstrom information $\|\hat{S}^\sigma\|_\rho^2 = \|S^\sigma\|_\rho^2$. Nevertheless, it is unimportant to make the distinction between the element and its observables in the following.

Let $\mathcal{S}$ be the set of all regular 1D submodels of $\mathbf{G}$. Let $\{S\}$ be the tangent set, defined as the set of the scores of all such submodels, viz.,

$$\{S\} \equiv \{S^\sigma : \sigma \in \mathcal{S}\} \subseteq \mathcal{Z}, \tag{3.8}$$

and let the tangent space $\mathcal{T}$ be the closed linear span of the tangent set, viz.,

$$\mathcal{T} \equiv \overline{\mathrm{span}}\{S\} \subseteq \mathcal{Z}. \tag{3.9}$$

Define also the set of influence observables with respect to β as

$$\mathcal{D} \equiv \{\delta \in \mathcal{Z} : \langle \delta, S^\sigma \rangle_\rho = \dot{\beta}^\sigma \, \forall \sigma \in \mathcal{S}\}, \tag{3.10}$$

where $\dot{\beta}^\sigma$ is obtained by expressing β as a function of θ for the submodel $\{\sigma(\theta)\}$ and taking the derivative. Let $\mathcal{E}$ be the positive-operator-valued measure that models a measurement. The mean-square error of an unbiased estimator $\check{\beta}$ at the truth becomes

$$\mathsf{E} \equiv \int [\check{\beta}(\lambda) - \beta]^2 \mathrm{tr}\mathcal{E}(d\lambda)\rho. \tag{3.11}$$

If $\mathcal{D}$ is not empty, the generalized Helstrom bound on E for any measurement and any unbiased estimator is given by

$$\mathsf{E} \geqslant \tilde{\mathsf{H}} = \|\delta_{\mathrm{eff}}\|_\rho^2, \tag{3.12}$$

where δ_{eff} is the efficient influence given by the orthogonal projection of any $\delta \in \mathcal{D}$ into $\mathcal{T}$, defined by

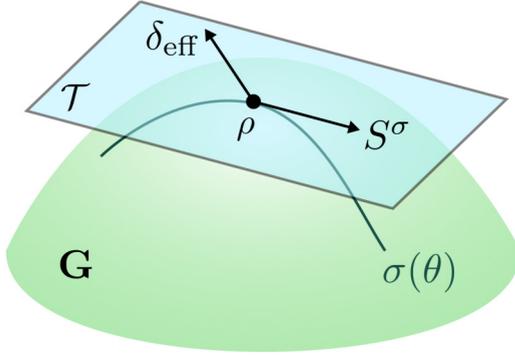$$\delta_{\mathrm{eff}} \equiv \Pi(\delta | \mathcal{T}) \, \forall \delta \in \mathcal{D}. \tag{3.13}$$

FIG. 2. Geometric picture of the family of states **G** as a manifold, the true state ρ as a point, a 1D submodel $\sigma(\theta)$ as a line, its score S^σ as a tangent vector, $\mathcal{T}$ as the tangent space, and the efficient influence δ_{eff} as another vector, which is the gradient of β in $\mathcal{T}$.

The result of the projection is unique, as $\mathcal{D}$ is the affine subspace $\delta_{\text{eff}} + \mathcal{T}^\perp$, where $\mathcal{T}^\perp$ is the orthocomplement of $\mathcal{T}$ in $\mathcal{Z}$. Figure 2 illustrates the essential concepts from the geometric perspective. As discussed in Ref. [36], the tangent space for the imaging problem turns out to be nontrivial and a closed-form solution for the generalized Helstrom bound is difficult to obtain, so I resort to looser bounds in the following. A useful lemma, generalizing the classical Lemma 25.19 in Ref. [43] and a similar result for the finite-dimensional quantum case in Ref. [44], is as follows.

Lemma 2. Let

$$\tilde{\mathsf{H}} = \sup_{u \in \text{span}\{S\}} \mathsf{H}(u), \tag{3.14}$$

$$\mathsf{H}(u) \equiv \|\Pi(\delta|\text{span } u)\|_\rho^2 \,\forall\, \delta \in \mathcal{D}, \tag{3.15}$$

where span$u$ is the Hilbert space spanned by one element u. If $u \neq 0$,

$$\mathsf{H}(u) = \frac{\langle \delta, u \rangle_\rho^2}{\|u\|_\rho^2}. \tag{3.16}$$

The proof is deferred to Appendix B.

If the element u in span$\{S\}$ is the score of a 1D submodel, as is often the case, $\mathsf{H}(u)$ is the Helstrom bound and $\langle \delta, u \rangle_\rho = \dot{\beta}$ with respect to the submodel. Lemma 2 then implies that one may consider only 1D submodels in the evaluation or bounding of $\tilde{\mathsf{H}}$. To obtain a tight lower bound on $\tilde{\mathsf{H}}$ via $\mathsf{H}(u)$, the submodel should be made as unfavorable to the estimation as possible.

IV. UNFAVORABLE PARAMETRIC SUBMODEL

For the imaging problem, let

$$\mathbf{G} = \left\{ [\rho(P)]^{\otimes M} : P \in \mathcal{P} \right\}, \tag{4.1}$$

$$\mathcal{P} = \text{all probability measures on } (\mathbb{R}, \Sigma), \tag{4.2}$$

where $\rho(P)$ is given by Eqs. (2.1)–(2.3). Equation (4.2) implies that no knowledge about the object is assumed, other than the fact that it is spatially incoherent and the expected total photon number N is known. If N is unknown, a submodel with a fixed N still gives a valid lower bound on $\tilde{\mathsf{H}}$, by
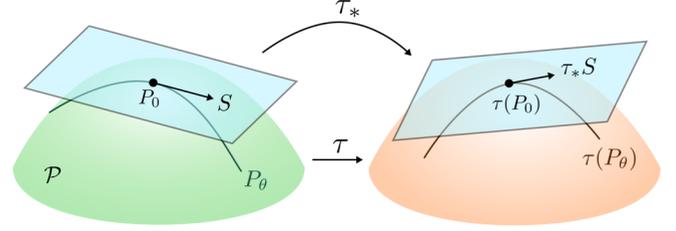

FIG. 3. Here τ is a map from a set of input probability measures to a set of output quantum states and τ_* is the pushforward of each score from the input tangent space to the output tangent space.

Lemma 2. Let the parameter of interest be the linear functional

$$\beta(P) = \int b(x)P(dx) \tag{4.3}$$

for a given real function $b(x)$. If $b(x)$ is a monomial, viz.,

$$b(x) = x^\mu, \quad \mu \in \mathbb{N}_1, \tag{4.4}$$

I call the β a simple moment, as considered in previous works [8–14]. If β is not necessarily simple, I call it a generalized moment.

Let $P_0 \in \mathcal{P}$ be the true measure and consider the submodel

$$P_\theta(u) = \int_u g_\theta(x)P_0(dx), \quad u \in \Sigma, \tag{4.5}$$

$$g_\theta(x) \equiv \frac{1 + \tanh[\theta S(x)]}{\int \{1 + \tanh[\theta S(x)]\}P_0(dx)}, \tag{4.6}$$

$$\theta \in (-c, c), \quad 0 < c < \infty, \tag{4.7}$$

where the Radon-Nikodym derivative $g_\theta(x)$ given by Eq. (4.6) is chosen for its convenient properties [45] (see also Lemma 8 in Appendix D) and $S(x)$ is assumed to be a function in $L_2(P_0)$ with zero mean at the truth, viz.,

$$\int S(x)P_0(dx) = \langle 1, S \rangle_{P_0} = 0. \tag{4.8}$$

Each P_θ is a valid probability measure and $\{P_\theta : \theta \in (-c, c)\}$ also contains the true P_0 at $\theta = 0$, so $\{P_\theta\}$ is a valid submodel for the purpose of Lemma 2.

It is straightforward to show that $S(x)$ is the classical score of the submodel $\{P_\theta\}$. The quantum score of the corresponding submodel $\{\tau(P_\theta) : \theta \in (-c, c)\}$ for each photon is the pushforward of S by the map τ,[1] denoted by τ_*S, to borrow the terminology from differential geometry [46]. Figure 3 illustrates the concept from the geometric perspective. Formally, the observables in the equivalence class of τ_*S obey the Lyapunov equation

$$(\tau_*S) \circ \tau(P_0) = \dot{\tau} = \tau(\dot{P}) = \tau(SP_0) \tag{4.9}$$

$$= \int e^{-i\hat{k}x} |\psi\rangle \langle\psi| e^{i\hat{k}x} S(x)P_0(dx). \tag{4.10}$$

---

[1]The τ_* is called the score operator in the classical statistics literature [43,45], but that term would cause confusion with the quantum operators so I do not use it.

The Helstrom bound for the M-temporal-mode submodel $\{[\rho(P_\theta)]^{\otimes M} : \theta \in (-c, c)\}$ becomes

$$\mathsf{H} = \frac{\dot{\beta}^2}{N \|\tau_* S\|^2_{\tau(P_0)}}, \tag{4.11}$$

$$\dot{\beta} = \int b(x) S(x) P_0(dx) = \langle b, S \rangle_{P_0}. \tag{4.12}$$

By Lemma 2, Eq. (4.11) is a lower bound on the generalized Helstrom bound $\check{\mathsf{H}}$ for the semiparametric problem. The prequel to this paper assumes a 1D model with $\theta = \beta$ being a simple moment of order μ while all the other moments are fixed [12]. It justifies the quantum bound by appealing to the inequalities $\mathsf{E} \geqslant (J^{-1})_{\mu\mu} \geqslant 1/J_{\mu\mu} \geqslant 1/K_{\mu\mu}$ for the Fisher information matrix J and the Helstrom information matrix K. However, it is unclear whether such a model is a valid submodel and whether the inequalities are justified for the infinite-dimensional semiparametric model. Reference [11], a related work that studies the classical Fisher information for the same moment estimation problem, shares the same issues. Here I am able to alleviate these doubts by explicitly constructing a valid submodel and appealing to Lemma 2.

Note that multiplying S by any nonzero constant does not change H, as $\tau_* S$ and $\langle b, S \rangle_{P_0}$ are both linear with respect to S, so there is no loss of generality if S is normalized to

$$\|S\|_{P_0} = 1. \tag{4.13}$$

To make the submodel bound H a tight lower bound on the semiparametric $\check{\mathsf{H}}$, there are two heuristic considerations in the choice of S: (i) The τ map should shrink its norm as much as possible and (ii) $\dot{\beta} = \langle b, S \rangle_{P_0} \neq 0$. Since τ models the imaging process with a bandwidth limit, physical intuition suggests that picking a highly oscillatory function for S may lead to significant norm shrinkage. A convenient choice in this regard is an orthonormal polynomial $a_n(x)$ specified by Lemma 4 in Appendix A, defined here with respect to the true measure P_0. It is already normalized and each $a_n(x)$ is more oscillatory for higher n, as each $a_n(x)$ has n zeros within the support of P_0 [47].

For the second consideration, suppose that β is a simple moment with respect to the $b(x)$ given by Eq. (4.4). Then the highest n for which $\langle a_n, b \rangle_{P_0} \neq 0$ is $n = \mu$ by Lemma 4 in Appendix A. Thus, a promising choice is

$$S(x) = a_\mu(x). \tag{4.14}$$

V. PURIFICATION BOUNDS

A closed-form solution for the Helstrom information of the mixed-state submodel remains difficult to obtain. A standard technique in quantum metrology is to bound it using a purified model and the monotonicity of Helstrom information [40,41,48].

Lemma 3. Let $\{\omega_\theta : \theta \in \mathcal{R} \subseteq \mathbb{R}\}$ be a regular model of states on $\mathcal{H} \otimes \mathcal{H}'$ and $\{\tau_\theta = \mathrm{tr}'\omega_\theta : \theta \in \mathcal{R}\}$ be the model on $\mathcal{H}$ generated by the partial trace. Then $\{\tau_\theta\}$ is also regular and

$$\|\mathrm{tr}'_* S\|^2_{\mathrm{tr}'\omega} \leqslant \|S\|^2_\omega, \tag{5.1}$$

where S is the score of $\{\omega_\theta\}$ at the true ω and $\mathrm{tr}'_* S$ is the score of $\{\tau_\theta\}$ at $\mathrm{tr}'\omega$.

The monotonicity is well established for finite-dimensional states [40] and commonly assumed in the physics literature even for infinite-dimensional states [41,48]. As I am unable to find a rigorous proof for the general case in the literature, I present one in Appendix C for completeness.

For the submodel $\{\tau(P_\theta)\}$ in Sec. IV, the natural purification given by Eq. (2.12) can be written as

$$\tau(P_\theta) = \mathrm{tr}'\Psi_\theta, \tag{5.2}$$

$$|\Psi_\theta\rangle = e^{-i\hat{k}\otimes\hat{x}} |\psi\rangle \otimes \sqrt{g_\theta(\hat{x})} |\phi_0\rangle, \tag{5.3}$$

where $|\phi_0\rangle$ is a purification of P_0. The score of $\{\Psi_\theta : \theta \in (-c, c)\}$ is then $\hat{S} = S(\hat{x})$ and Lemma 3 gives

$$\|\tau_* S\|^2_{\tau(P_0)} = \|\mathrm{tr}'_*\hat{S}\|^2_{\mathrm{tr}'\Psi_0} \leqslant \|\hat{S}\|^2_{\Psi_0} = \|S\|^2_{P_0} = 1, \tag{5.4}$$

which is a loose bound that does not depend on $|\psi\rangle$ of the imaging system. I therefore turn to the alternative purification in Lemma 1. Each P_θ given by Eq. (4.5) satisfies the condition of infinite and bounded support for Lemma 1 as long as the true P_0 satisfies it, as each P_θ is dominated by P_0 and the $g_\theta(x)$ in Eq. (4.6) is strictly positive. I can then use Lemma 1 to write

$$\tau(P_\theta) = \mathrm{tr}''\Phi_\theta, \tag{5.5}$$

$$|\Phi_\theta\rangle = \sum_{p=0}^{\infty} \sum_{n=0}^{p} \frac{(-i\hat{k})^p}{p!} L_{pn}(\theta) |\psi\rangle \otimes |n\rangle, \tag{5.6}$$

where the parameter dependence comes from the Cholesky factor $L(\theta)$ of the Hankel matrix

$$H_{pq}(\theta) = \langle x^p, x^q \rangle_{P_\theta} = \int x^{p+q} P_\theta(dx). \tag{5.7}$$

Note that the isometry used to generate this purification depends on P and therefore θ for the submodel, so the resulting Helstrom information may differ from that using the original purification.

Let S_Φ be the score of $\{\Phi_\theta : \theta \in (-c, c)\}$. Then, by Lemma 3,

$$\|\tau_* S\|^2_{\tau(P_0)} = \|\mathrm{tr}''_* S_\Phi\|^2_{\mathrm{tr}''\Phi_0} \leqslant \|S_\Phi\|^2_{\Phi_0}. \tag{5.8}$$

Figure 4 illustrates the concepts introduced thus far from the geometric perspective. While a closed-form solution for $\|S_\Phi\|^2_{\Phi_0}$ is still intractable, its scaling with the true object size Δ defined by Eq. (2.17) in the asymptotic limit $\Delta \to 0$ can be proved. Without loss of generality, Δ can be measured in dimensionless "Airy units" relative to an effective bandwidth of the imaging system. Then $\Delta \ll 1$ defines the regime of subdiffraction objects. In the following, the asymptotic notation of $O(u(\Delta))$ [order at most $u(\Delta)$], $o(u(\Delta))$ [order smaller than $u(\Delta)$], $\Omega(u(\Delta))$ [order at least $u(\Delta)$], and $\Theta[u(\Delta)]$ [order exactly $u(\Delta)$] will be used [49,50].

Theorem 1. Assume the semiparametric model given by Eqs. (2.1)–(2.3), (4.1), and (4.2). Assume that $\hat{k}$ is bounded. Assume further that P_0 is in the Szegő class [51], viz., P_0 has a density $f_0(x)$ with respect to the Lebesgue measure on a compact interval $[c_1, c_2]$, $c_1 < c_2$, such that

$$\int_{c_1}^{c_2} \frac{\ln f_0(x)}{\sqrt{(c_2 - x)(x - c_1)}} dx > -\infty. \tag{5.9}$$
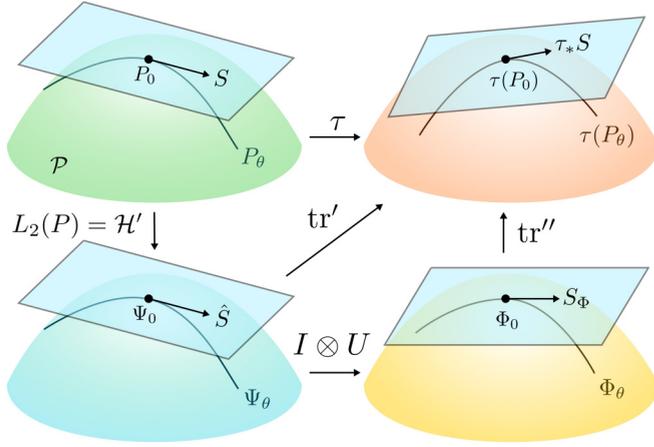
FIG. 4. Geometric picture that summarizes all the purifications, parametric submodels, scores, and maps that are involved in the proof of Theorem 1. With the maps τ, tr', and tr'' all leading to the same model $\{\tau(P_\theta)\}$, the final score can be expressed as the pushforwards $\tau_* S = \mathrm{tr}'_* \hat{S} = \mathrm{tr}''_* S_\Phi$.

For the parameter of interest given by Eq. (4.3) and

$$b(x) = x^\mu + o(\Delta^\mu), \quad \mu \in \mathbb{N}_1, \tag{5.10}$$

the generalized Helstrom bound is

$$\tilde{\mathsf{H}} = \frac{\Omega(\Delta^{2\lfloor \mu/2 \rfloor})}{N}. \tag{5.11}$$

The proof is deferred to Appendix D.

While Theorem 1 appears to be identical to the main result in Ref. [12], the rigor has been much improved here. Most notably, (i) Lemma 2 ensures that the 1D submodel bound is valid for an infinite-dimensional model, avoiding the use of the questionable parametric model and matrix algebra in Refs. [11,12], (ii) Lemma 3 ensures that the monotonicity of the Helstrom information remains true for infinite-dimensional states, and (iii) the conditions for Theorem 1 have been rigorously established, avoiding the questionable algebra of possibly unbounded operators in Ref. [12].

The important physical implication of Theorem 1 is that the mean-square error E of any unbiased estimator must decrease with Δ more slowly than the signal

$$\beta^2 = O(\Delta^{2\mu}), \tag{5.12}$$

so the signal-to-noise ratio is

$$\frac{\beta^2}{\mathsf{E}} = N O(\Delta^{2\lceil \mu/2 \rceil}) \tag{5.13}$$

and the moments of smaller subdiffraction objects are harder to estimate, especially for higher orders.

Notice that Eq. (5.10) generalizes the theorem for a larger class of parameters, beyond the simple moments considered in previous works [8–12]. I call a moment associated with Eq. (5.10) a generalized moment of order μ. The theorem shows that, regardless of the $o(\Delta^\mu)$ terms, the quantum limits for generalized moments of the same order have the same leading-order term. This fact will be useful in the proof of the optimality of SPADE in Sec. VI.

The formalism here may also be able to deal with a $\beta(P)$ that is nonlinear with respect to P, such as the entropy, in which case $b(x)$ should be replaced by a gradient of $\beta(P)$ in the $L_2(P_0)$ space, although this case is outside the scope of the paper.

VI. SPATIAL-MODE DEMULTIPLEXING WITH ONE OR TWO MODES

Suppose that the spatial-frequency measure Q, defined in Eq. (2.8) with respect to $\hat{k}$ and $|\psi\rangle$ of the imaging system, has an infinite and bounded support. Then the orthonormal sequence

$$\{|\psi_n\rangle \equiv (-i)^n \tilde{a}_n(\hat{k}) |\psi\rangle : n \in \mathbb{N}_0\}, \tag{6.1}$$

defined in terms of the orthonormal polynomials $\{\tilde{a}_n\}$ for the measure Q (as per Lemma 4 in Appendix A), can be used to construct measurements of the light. The set is called the point-spread-function-adapted (PAD) modes [9,52], generalizing the Hermite-Gaussian modes for a Gaussian Q.

It is interesting to note that the orthonormal polynomials $\{a_n\}$ for P play central roles in previous sections, enabling the purification given by Lemma 1 (as shown in Appendix A) and also providing the score for the submodel used in Theorem 1. Following similar steps, it can be shown that the probability of each photon being projected into a PAD mode is

$$q_n(P) \equiv \langle \psi_n | \tau(P) | \psi_n \rangle = \int |C_n(x)|^2 P(dx), \tag{6.2}$$

$$C_n(x) \equiv \langle \psi_n | e^{-i\hat{k}x} | \psi \rangle = \sum_{p=0}^\infty \frac{(-ix)^p i^n}{p!} \tilde{L}_{pn}, \tag{6.3}$$

where $\tilde{L}_{pn}$ is now the Cholesky factor of the Hankel matrix with respect to Q. A property of $\tilde{L}_{pn}$ is that it is zero for $p < n$ and nonzero for $p = n$ (as per Lemma 4 in Appendix A), so

$$C_n(x) = \frac{\tilde{L}_{nn}}{n!} x^n + O(\Delta^{n+1}), \tag{6.4}$$

$$q_n(P) = r_n \beta_{2n}(P), \quad r_n \equiv \frac{\tilde{L}_{nn}^2}{n!^2}, \tag{6.5}$$

$$\beta_{2n}(P) = \int [x^{2n} + O(\Delta^{2n+1})] P(dx). \tag{6.6}$$

In other words, each probability q_n is proportional to a certain generalized object moment β_{2n} of even order $\mu = 2n$.

If Q is Gaussian, say,

$$Q(dk) = \sqrt{\frac{2}{\pi}} e^{-2k^2} dk, \tag{6.7}$$

then its support is unbounded, but the preceding discussion still holds, as Eq. (6.3) has the closed-form solution [8]

$$C_n(x) = e^{-x^2/8} \frac{(x/2)^n}{\sqrt{n!}}. \tag{6.8}$$

There may exist more general conditions on Q for the results in this section to be valid, but such a generalization does not seem to be interesting from the physics perspective and is therefore not pursued in this work.

Let $\{\mathcal{N}_n : n \in \mathbb{N}_0\}$ be the integrated photon counts from the PAD-mode projections over the M temporal modes. The

expected value of each count is

$$\mathbb{E}(\mathcal{N}_n) = M\epsilon q_n = N r_n \beta_{2n}. \tag{6.9}$$

Assuming a known N (the theory for an unknown N is similar [9,10]), an unbiased estimator of $\beta_{2n}(P)$ is therefore

$$\check{\beta}_{2n} = \frac{\mathcal{N}_n}{r_n N}. \tag{6.10}$$

To estimate an odd generalized moment, consider the so-called interferometric PAD (IPAD) modes [8,9]

$$|\psi_n^+\rangle \equiv \frac{|\psi_n\rangle + |\psi_{n+1}\rangle}{\sqrt{2}}, \quad |\psi_n^-\rangle \equiv \frac{|\psi_n\rangle - |\psi_{n+1}\rangle}{\sqrt{2}}. \tag{6.11}$$

Let the integrated photon counts resulting from projections in this pair of modes be $\mathcal{N}_n^+$ and $\mathcal{N}_n^-$. Proceeding in the same way as before, it can be shown that

$$\mathbb{E}(\mathcal{N}_n^+) = \frac{N}{2}[r_n \beta_{2n} + s_n \beta_{2n+1} + O(\Delta^{2n+2})], \tag{6.12}$$

$$\mathbb{E}(\mathcal{N}_n^-) = \frac{N}{2}[r_n \beta_{2n} - s_n \beta_{2n+1} + O(\Delta^{2n+2})], \tag{6.13}$$

where $s_n \equiv 2\tilde{L}_{nn}\tilde{L}_{n+1\,n+1}/[n!(n+1)!]$ and

$$\beta_{2n+1}(P) = \int [x^{2n+1} + O(\Delta^{2n+2})]P(dx) \tag{6.14}$$

is a certain odd generalized moment. An unbiased estimator of $\beta_{2n+1}(P)$ is therefore

$$\check{\beta}_{2n+1} = \frac{\mathcal{N}_n^+ - \mathcal{N}_n^-}{s_n N}. \tag{6.15}$$

See Refs. [8,9] for further details about how bases may be constructed from the PAD and IPAD modes.

The following proposition summarizes the SPADE performance.

Proposition 1. Assume that the spatial-frequency measure Q either has an infinite and bounded support or is Gaussian. Then projections into the PAD modes enable unbiased estimation of a certain set of even generalized moments $\{\beta_{2n} : n \in \mathbb{N}_0\}$, with the variance of each estimator $\check{\beta}_{2n}$ given by

$$\mathbb{V}(\check{\beta}_{2n}) = \frac{O(\Delta^{2n})}{N}[1 + O(\epsilon)], \tag{6.16}$$

while projections into the IPAD modes enable unbiased estimation of a certain set of odd generalized moments $\{\beta_{2n+1} : n \in \mathbb{N}_0\}$, with the variance of each estimator $\check{\beta}_{2n+1}$ given by

$$\mathbb{V}(\check{\beta}_{2n+1}) = \frac{O(\Delta^{2n})}{N}[1 + O(\epsilon)]. \tag{6.17}$$

The variances achieve the quantum-limited Δ scalings given by Eq. (5.11) in Theorem 1.

Proof. Given Eq. (2.1), $\mathcal{N}$ is a binomial process [53], which leads to

$$\mathbb{V}(\mathcal{N}_n) = N q_n (1 - \epsilon q_n) = N r_n \beta_{2n}[1 + O(\epsilon)]. \tag{6.18}$$

The variance of the estimator given by Eq. (6.10) is then

$$\mathbb{V}(\check{\beta}_{2n}) = \frac{\mathbb{V}(\mathcal{N}_n)}{r_n^2 N^2} = \frac{\beta_{2n}}{r_n N}[1 + O(\epsilon)], \tag{6.19}$$

which gives Eq. (6.16), since $\beta_{2n} = O(\Delta^{2n})$. The derivation of Eq. (6.17) using the basic properties of a binomial process is similar. ∎

Note that the conditions for Proposition 1 are more relaxed than those for Theorem 1.

The important point is that, although a finite number of PAD or IPAD modes cannot measure a simple moment of the object exactly, they can provide exact unbiased estimators of the moments they are naturally measuring. By considering the latter as the parameters of interest, Proposition 1 proves the quantum optimality of SPADE with just one or two modes, at least in terms of the Δ scalings.

VII. BOUND FOR DIRECT IMAGING

The semiclassical model of direct imaging is to assume a probability density $\eta(\xi, P)$ for each photon with respect to the Lebesgue measure that obeys

$$\eta(\xi, P) = \int h(\xi - x)P(dx), \tag{7.1}$$

where h is the non-negative point-spread function for direct incoherent imaging. For a diffraction-limited system, it is given by

$$h(\xi) = \left|\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ikx}\langle k|\psi\rangle\,dk\right|^2, \tag{7.2}$$

where $\langle k|\psi\rangle$ is the optical transfer function [2]. Let $\nu^{(M,\epsilon)}[P]$ denote the probability measure for the photon counts over M temporal modes. Previous studies [8–10] have suggested that, for the semiparametric model $\{\nu^{(M,\epsilon)}[P] : P \in \mathcal{P}\}$ based on Eq. (4.2), the Cramér-Rao bound for the estimation of a simple moment with direct imaging is

$$\mathsf{E} \geqslant \tilde{\mathsf{C}} = \frac{\Theta(1)}{N}. \tag{7.3}$$

Unfortunately, this result is rigorously proven only when the location family $\{h(\xi - x) : x \in \text{supp}P_0\}$ satisfies a special statistical property called completeness [54] (see Ref. [45], Sec. 6.5; Ref. [43], Sec. 25.5.2; and Ref. [10]). While a Gaussian h satisfies the property, completeness turns out to be hard to prove more generally, and Eq. (7.3) remains questionable for any non-Gaussian h.

I turn again to the parametric-submodel approach in Sec. IV, which leads to the following proposition.

Proposition 2. Assume that the derivatives $h^{(n)}(\xi)$ of the point-spread function $h(\xi)$ for direct imaging exist, up to order $n = \mu + 1$. If there exist non-negative x-independent functions $\underline{h}(\xi)$ and $\bar{h}(\xi)$ such that, for all $\xi \in \mathbb{R}$ and for all $x \in \text{supp}P_0$,

$$h(\xi - x) \geqslant \underline{h}(\xi), \tag{7.4}$$

$$|h^{(\mu+1)}(\xi - x)| \leqslant \bar{h}(\xi), \tag{7.5}$$

and the functions satisfy

$$\int_{-\infty}^{\infty} \frac{[h^{(\mu)}(\xi)]^2}{\underline{h}(\xi)}d\xi < \infty, \quad \int_{-\infty}^{\infty} \frac{[\bar{h}(\xi)]^2}{\underline{h}(\xi)}d\xi < \infty, \tag{7.6}$$

then the Cramér-Rao bound for the estimation of a generalized moment of order μ is

$$\tilde{C} = \frac{\Omega(1)}{N}. \tag{7.7}$$

The proof is deferred to Appendix E.

The conditions for Proposition 2 seem specific, but they can be checked for simple functions such as

$$h(\xi) = \frac{d_1}{(\xi/d_2)^{2p} + 1}, \quad h(\xi) = d_1 e^{-(\xi/d_2)^{2p}} \tag{7.8}$$

for some positive constants d_1 and d_2 and positive integer p. The tails of $h^{(p)}(\xi - x)$ for these functions decay at least as fast as the tails of $h(\xi)$, while an $\underline{h}(\xi)$ to lower bound $h(\xi - x)$ can be obtained by replacing ξ^2 by $(|\xi| + \Delta_0)^2$ in Eqs. (7.8) for some Δ_0 that upper bounds all Δ of interest, as that would give $|\xi - x| \leqslant |\xi| + |x| \leqslant |\xi| + \Delta \leqslant |\xi| + \Delta_0$ and a lower bound on $h(\xi - x)$ for all $x \in \mathrm{supp} P_0$. Then $\underline{h}(\xi)$ remains positive and has tails that decay in the same way as $h(\xi)$, leading to the finite integrals in Eqs. (7.6). The advantage of Proposition 2 over the previous approaches in Refs. [8–10] is that the conditions for the former are still easier to check than proving the completeness property and Proposition 2 can remain valid even for an incomplete location family.

The physical implication of Proposition 2 is that, for a generalized moment of order $\mu \geqslant 2$, both the quantum limit given by Theorem 1 and the SPADE performance given by Proposition 1 are substantial improvements over direct imaging. Nevertheless, there exist counterexamples in which the point-spread function has zeros and the conditions for Proposition 2 are violated. The argument by Paúr *et al.* [55,56] in particular suggests that the zeros can enhance the Fisher-information integral for the submodel by a $\Theta(1/\Delta)$ factor, leading to

$$\tilde{C} = \frac{\Omega(\Delta)}{N}, \tag{7.9}$$

but despite some numerical evidence (unpublished), I am unable to prove this bound in a general fashion. It therefore remains an open problem whether Eq. (7.9) is the ultimate limit to direct imaging or by how much the zeros in its point-spread function can improve it.

VIII. CONCLUSION

The key results of this work are the quantum limit given by Theorem 1, the SPADE performance given by Proposition 1, and the direct-imaging bound given by Proposition 2. They confirm rigorously the prior intuition that the moments of smaller subdiffraction objects are harder to estimate, especially for higher orders, although SPADE can estimate them with optimal error scalings, while direct imaging is unlikely to be nearly as efficient. Beyond the improved rigor, a useful advance is the generalization of the results for a larger class of moments such that the quantum optimality of SPADE with a finite number of modes is proved and its experimental demonstration becomes much easier.

The price to pay for the generality of the results here is their imprecise nature in terms of the asymptotic notions. More precise results can be obtained numerically for more special cases, as has been done in Refs. [8–10] regarding the

SPADE and direct-imaging performances. To compute concrete quantum and classical limits, the parametric-submodel approach should help, as it is able to give bounds for an infinite-dimensional model through 1D submodels. The submodel bounds can be computed numerically without resorting to purifications, at least for special cases of the true measure P_0.

Other interesting future directions include more rigorous proofs of the convergence of thermal models to the binomial and Poisson models considered here, the study of more general types of parameters, Bayesian and minimax approaches [57–59], and—of course—experimental demonstrations of quantum-limited moment estimation. Beyond imaging, the model and the results here may be applied or generalized to other sensing applications, such as the estimation of diffusion parameters in phase estimation and optomechanics [60,61]. The general principles established in this work thus give rigorous underpinnings to the foundations of imaging theory and possibly beyond.

APPENDIX A: PROOF OF LEMMA 1

Before proving Lemma 1, I collect some basic facts in the following lemma.

Lemma 4. In this lemma, the sample space is $\mathbb{R}$, the inner product is given by Eq. (2.4), and the measure P has finite moments and $\#\mathrm{supp} P = J$, where J may be infinite. Let $A_{(j)}$ denote the $(j + 1) \times (j + 1)$ upper left submatrix of a matrix A for a $j \in \mathbb{N}_{[0,J)} \equiv \{j \in \mathbb{N}_0 : j < J\}$.

(i) The $H_{(j)}$ for the Hankel matrix given by Eq. (2.19) is positive definite ($H_{(j)} > 0$) for any $j \in \mathbb{N}_{[0,J)}$.

(ii) The monomials $\{x^p : p \in \mathbb{N}_{[0,J)}\}$ are linearly independent.

(iii) Define the orthonormal polynomials $\{a_n(x) : n \in \mathbb{N}_{[0,J)}\}$ as

$$a_n(x) = \sum_{p=0}^{n} A_{np} x^p, \tag{A1}$$

where A is the lower triangular matrix ($A_{np} = 0$ if $p > n$) obtained by applying the Gram-Schmidt procedure to the monomials, such that

$$\langle a_n, a_m \rangle_P = \delta_{nm}. \tag{A2}$$

One can make $A_{nn} > 0$ for all n. The orthonormality implies, for any $j \in \mathbb{N}_{[0,J)}$,

$$A_{(j)} H_{(j)} A_{(j)}^{\top} = I_{(j)}, \quad H_{(j)}^{-1} = A_{(j)}^{\top} A_{(j)}, \tag{A3}$$

where $\top$ denotes the transpose and I is the identity matrix.

(iv) Let L be the lower triangular Cholesky factor of H. Then, for any $j, n, p \in \mathbb{N}_{[0,J)}$,

$$L_{(j)} L_{(j)}^{\top} = H_{(j)}, \quad L_{(j)} = A_{(j)}^{-1}, \tag{A4}$$

$$L_{nn} = \frac{1}{A_{nn}} > 0, \tag{A5}$$

$$x^p = \sum_{n=0}^{p} L_{pn} a_n(x), \quad \langle x^p, a_n \rangle_P = L_{pn}. \tag{A6}$$

Proof of Lemma 4. I prove only statement (i) here; the rest is basic linear algebra (see, for example, Ref. [47]). Write $C = \text{supp} P$ for brevity in this proof. Recall that the support is defined uniquely by the following conditions (see [32], Theorem 2.1): (i) $P(C) = 1$ and (ii) if a closed set D satisfies $P(D) = 1$, then $C \subseteq D$. In other words, the support is the smallest closed set with unit probability. The second condition is equivalent to the condition that any closed strict subset D of C must give $P(D) < 1$.

Given any column vector $u \in \mathbb{R}^{j+1}$ with $u^\top u \neq 0$, let

$$\tilde{u}(x) \equiv \sum_{p=0}^{j} u_p x^p, \quad Z \equiv \{x \in C : \tilde{u}(x) = 0\}. \tag{A7}$$

Then

$$[\tilde{u}(x)]^2 > 0 \,\forall\, x \in C - Z. \tag{A8}$$

As any nonzero polynomial of degree j has at most j isolated zeros, $\#Z \leqslant j$ and Z is closed. Since $\#C = J > j \geqslant \#Z$, $C - Z$ cannot be empty, Z is a strict subset of C, and by the definition of C,

$$P(Z) < 1, \quad P(C - Z) = P(C) - P(Z) > 0. \tag{A9}$$

Now consider

$$u^\top H_{(j)} u = \int_C [\tilde{u}(x)]^2 P(dx) = \int_{C-Z} [\tilde{u}(x)]^2 P(dx). \tag{A10}$$

I wish to prove that the last expression is not zero by contradiction. Suppose that it is zero. Then, by Proposition 4.1.7 in Ref. [31],

$$[\tilde{u}(x)]^2 = 0 \quad \text{a.e. on } C - Z, \tag{A11}$$

where a.e. denotes almost everywhere. Equation (A11) implies the existence of a set Y with $P(Y) = 0$ such that $[\tilde{u}(x)]^2 = 0$ for all $x \in C - Z - Y$. Here $C - Z - Y$ is not empty, because $P(C - Z - Y) = P(C - Z) > 0$. Then, for any $x \in C - Z - Y \subseteq C - Z$, the $[\tilde{u}(x)]^2 = 0$ statement contradicts Eq. (A8). By contradiction, Eq. (A10) cannot be zero and must be strictly positive. As u is arbitrary, $H_{(j)} > 0$. ∎

Proof of Lemma 1. If the support of P is infinite and bounded, the set $\{a_n(x) : n \in \mathbb{N}_0\}$ defined in Lemma 4 is an orthonormal basis of $L_2(P)$ (see [31], Proposition 6.4.1). Let its $\mathcal{H}'$ representation be

$$a_n(x) \leftrightarrow a_n(\hat{x}) |\phi\rangle \equiv |\varphi_n\rangle. \tag{A12}$$

Then $\{|\varphi_n\rangle : n \in \mathbb{N}_0\}$ is the same orthonormal basis in the new representation and a resolution of the identity $I_{\mathcal{H}'}$ in the strong operator topology (see [30], Sec. 1.6) is

$$I_{\mathcal{H}'} = \sum_{n=0}^{\infty} |\varphi_n\rangle \langle \varphi_n|. \tag{A13}$$

An isometry $U : \mathcal{H}' \to \mathcal{H}''$ can then be written as

$$U = \sum_{n=0}^{\infty} |n\rangle \langle \varphi_n|, \tag{A14}$$

also in the strong operator topology. A proof is as follows. Let

$$I_j \equiv \sum_{n=0}^{j} |\varphi_n\rangle \langle \varphi_n|, \quad U_j \equiv \sum_{n=0}^{j} |n\rangle \langle \varphi_n|. \tag{A15}$$

Given any $|u\rangle \in \mathcal{H}'$ and $J > j$, consider

$$\|U_J |u\rangle - U_j |u\rangle\|^2 = \sum_{n=j+1}^{J} |\langle \varphi_n | u \rangle|^2 = s_J - s_j, \tag{A16}$$

where $\| |\cdot\rangle \| \equiv \sqrt{\langle \cdot | \cdot \rangle}$ and $s_j \equiv \sum_{n=0}^{j} |\langle \varphi_n | u \rangle|^2$. $\{s_j\}$ is monotonic, and since $\{|\varphi_n\rangle\}$ is an orthonormal sequence, $\{s_j\}$ is bounded by Bessel's inequality [see [25], Eq. (3.24)] and therefore convergent (see [24], Theorem 3.14) and Cauchy (see [24], Theorem 3.11). With the Cauchy property, given any $\varepsilon > 0$, there exists a j_0 such that, for all $J \geqslant j \geqslant j_0$, $s_J - s_j < \varepsilon^2$ and $\|U_J |u\rangle - U_j |u\rangle\| = \sqrt{s_J - s_j} < \varepsilon$. Thus, $\{U_j |u\rangle\}$ is also Cauchy. Since a Hilbert space is by definition complete (see [25], Definition 3.3.1), $\{U_j |u\rangle\}$ converges (see [25], Definition 1.4.5), and since $|u\rangle$ is arbitrary, $\{U_j\}$ converges strongly. By the same argument, $\{I_j\}$ and $\{U_j^\dagger\}$ also converge strongly. The strong convergences imply that $\{U_j^\dagger U_j = I_j\}$ converges strongly to $U^\dagger U = I_{\mathcal{H}'}$.

If the support of P is bounded, $\hat{x}$ can be assumed to be a bounded operator with a finite operator norm $\|\hat{x}\|_{\text{op}}$, since $\|\hat{x} |u\rangle\| = \|xu\|_P \leqslant \Delta \|u\|_P = \Delta \| |u\rangle \|$ for any $u \leftrightarrow |u\rangle$. As $\hat{k}$ is also assumed to be bounded, the exponential operator in Eq. (2.12) can be expressed as

$$e^{-i\hat{k} \otimes \hat{x}} = \sum_{p=0}^{\infty} \frac{(-i\hat{k})^p \otimes \hat{x}^p}{p!} \tag{A17}$$

in the sense of operator-norm convergence (see [29], Sec. VIII.4), which implies strong convergence. Let

$$F_j \equiv \sum_{p=0}^{j} \frac{(-i\hat{k})^p \otimes \hat{x}^p}{p!}. \tag{A18}$$

As $\{U_j\}$ and $\{F_j\}$ converge strongly, $\{(I_{\mathcal{H}} \otimes U_j)F_j\}$ also converges strongly. With the aid of Lemma 4, I obtain

$$\hat{x}^p |\phi\rangle = \sum_{n=0}^{p} L_{pn} a_n(\hat{x}) |\phi\rangle = \sum_{n=0}^{p} L_{pn} |\varphi_n\rangle, \tag{A19}$$

$$|\Phi^j\rangle \equiv (I_{\mathcal{H}} \otimes U_j) F_j |\psi\rangle \otimes |\phi\rangle \tag{A20}$$

$$= \sum_{p=0}^{j} \sum_{n=0}^{p} \frac{(-i\hat{k})^p}{p!} L_{pn} |\psi\rangle \otimes |n\rangle. \tag{A21}$$

The strong convergence of $\{(I_{\mathcal{H}} \otimes U_j)F_j\}$ implies the strong convergence of $\{|\Phi^j\rangle\}$, leading to Eq. (2.18). ∎

Remark. Throughout this paper, the convergence of a sequence of Hilbert-space elements is always assumed to be strong.

APPENDIX B: PROOF OF LEMMA 2

Proof of Lemma 2. Equation (3.16) comes from the fact that $\{e_1 \equiv u/\|u\|_\rho\}$ is the orthonormal basis of span$u$ and

$$\Pi(\delta|\text{span} u) = \langle \delta, e_1 \rangle_\rho e_1 = \frac{\langle \delta, u \rangle_\rho}{\|u\|_\rho^2} u. \quad \text{(B1)}$$

Of course, $\Pi(\delta|\text{span} 0) = 0$.

Now consider the supremum in Eq. (3.14), which is defined by two conditions [24]: (i) $\tilde{\mathsf{H}}$ is an upper bound on the set $\{\mathsf{H}(u) : u \in \text{span}\{S\}\}$ and (ii) $\tilde{\mathsf{H}} - \varepsilon$ is not an upper bound for any $\varepsilon > 0$. Since $\delta_{\text{eff}} \in \mathcal{D}$ [36], $\mathsf{H}(u)$ can also be expressed as

$$\mathsf{H}(u) = \|\Pi(\delta_{\text{eff}}|\text{span } u)\|_\rho^2. \quad \text{(B2)}$$

Then

$$\tilde{\mathsf{H}} = \|\delta_{\text{eff}}\|_\rho^2 \geqslant \|\Pi(\delta_{\text{eff}}|\text{span} u)\|_\rho^2 = \mathsf{H}(u) \, \forall u, \quad \text{(B3)}$$

by the shrinking property of the projection [25]. The first condition is therefore satisfied.

Assume now $\delta_{\text{eff}} \neq 0$, for $\tilde{\mathsf{H}} = \mathsf{H}(u) = 0 \; \forall u$ otherwise, and the lemma is trivial. As $\delta_{\text{eff}} \in \mathcal{T}$ and $\mathcal{T}$ is the closure of span$\{S\}$, there exists a sequence $\{u_j\} \subseteq \text{span}\{S\}$ that converges to δ_{eff} (see [25], Theorem 1.3.23). The convergence means that $\|u_j\|_\rho \to \|\delta_{\text{eff}}\|_\rho$ and $\langle u, u_j \rangle_\rho \to \langle u, \delta_{\text{eff}} \rangle_\rho$ for any $u \in \mathcal{T}$ (see [25], Theorem 3.3.12). With $\|\delta_{\text{eff}}\|_\rho > 0$, I can assume $\|u_j\|_\rho > 0$ for all $j \geqslant J$ and a sufficiently large J. Thus,

$$\mathsf{H}(u_j) = \frac{\langle \delta_{\text{eff}}, u_j \rangle_\rho^2}{\|u_j\|_\rho^2} \to \|\delta_{\text{eff}}\|_\rho^2 = \tilde{\mathsf{H}}. \quad \text{(B4)}$$

This limit implies that, given any $\varepsilon > 0$, there exists a $u \in \{u_j\} \subseteq \text{span}\{S\}$ such that

$$|\tilde{\mathsf{H}} - \mathsf{H}(u)| = \tilde{\mathsf{H}} - \mathsf{H}(u) < \varepsilon, \quad \mathsf{H}(u) > \tilde{\mathsf{H}} - \varepsilon. \quad \text{(B5)}$$

Thus, the second condition is also satisfied. ∎

APPENDIX C: PROOF OF LEMMA 3

Before proving Lemma 3, I present a couple of needed lemmas. The first is a trivial result regarding the partial trace and the L_2 spaces related by it. I write $I_{\mathcal{H}'} = I$ for brevity in this Appendix.

Lemma 5. Let ω be a state on $\mathcal{H} \otimes \mathcal{H}'$ and $\tau = \text{tr}'\omega$. Then

$$\langle u, v \rangle_\tau = \langle u \otimes I, v \otimes I \rangle_\omega \, \forall u, v \in L_2(\tau), \quad \text{(C1)}$$

$$\|u\|_\tau = \|u \otimes I\|_\omega \, \forall u \in L_2(\tau). \quad \text{(C2)}$$

Proof of Lemma 5. By the definition of partial trace (see [33], Proposition 16.6),

$$\text{tr} u \tau = \text{tr}(u \otimes I) \omega \, \forall u \in \mathcal{B}(\mathcal{H}). \quad \text{(C3)}$$

Then, for all $u, v \in \mathcal{B}(\mathcal{H})$,

$$\langle u, v \rangle_\tau = \text{tr}(u \circ v) \tau = \text{tr}[(u \otimes I) \circ (v \otimes I)] \omega$$
$$= \langle u \otimes I, v \otimes I \rangle_\omega. \quad \text{(C4)}$$

If any of $u, v \in L_2(\tau)$ are unbounded operators, let $\{u_j\}$ and $\{v_j\}$ be sequences in the dense subset $\mathcal{B}(\mathcal{H})$ that converge to them. Then (see [25], Theorem 3.3.12)

$$\langle u_j, v_j \rangle_\tau \to \langle u, v \rangle_\tau, \quad \text{(C5)}$$

$$\langle u_j \otimes I, v_j \otimes I \rangle_\omega \to \langle u \otimes I, v \otimes I \rangle_\omega. \quad \text{(C6)}$$

As Eq. (C4) holds for all $u_j, v_j \in \mathcal{B}(\mathcal{H})$, the right-hand sides of Eqs. (C5) and (C6) are also equal, proving Eq. (C1). Equation (C2) is a direct consequence of Eq. (C1). ∎

I now present a useful lemma regarding a quantum generalization of the conditional expectation, which appears in many other contexts [40,62,63]. The proof follows that of Theorem 6.1 in Ref. [40] for the finite-dimensional case.

Lemma 6. Let ω be a state on $\mathcal{H} \otimes \mathcal{H}'$ and $\tau = \text{tr}'\omega$. For each $X \in L_2(\omega)$, $\langle u \otimes I, X \rangle_\omega$ is a bounded linear functional of $u \in L_2(\tau)$. Define $\pi(X) \in L_2(\tau)$ as the unique Riesz representation that obeys

$$\langle u, \pi(X) \rangle_\tau = \langle u \otimes I, X \rangle_\omega \, \forall u \in L_2(\tau). \quad \text{(C7)}$$

Then the linear map $\Pi : L_2(\omega) \to L_2(\omega)$ defined as

$$\Pi X \equiv \pi(X) \otimes I \quad \text{(C8)}$$

is a projection. In particular, π satisfies the shrinking property

$$\|\pi(X)\|_\tau = \|\Pi X\|_\omega \leqslant \|X\|_\omega \, \forall X \in L_2(\omega). \quad \text{(C9)}$$

∎

Proof of Lemma 6. First note that $\langle u \otimes I, X \rangle_\omega$ is bilinear with respect to $u \in L_2(\tau)$ and $X \in L_2(\omega)$. For each $X \in L_2(\omega)$,

$$|\langle u \otimes I, X \rangle_\omega| \leqslant \|u \otimes I\|_\omega \|X\|_\omega = \|u\|_\tau \|X\|_\omega \quad \text{(C10)}$$

by the Cauchy-Schwarz inequality (CSI) and Lemma 5, so $\langle u \otimes I, X \rangle_\omega$ is a bounded linear functional of u and the Riesz representation theorem applies. Moreover, the theorem gives

$$\|\pi(X)\|_\tau = \|\pi\|_{\text{op}} = \sup_{\|u\|_\tau = 1} |\langle u \otimes I, X \rangle_\omega| \leqslant \|X\|_\omega, \quad \text{(C11)}$$

so $\pi : L_2(\omega) \to L_2(\tau)$ is a bounded linear map.

Now consider the Π map given by Eq. (C8). It is linear because π is linear. It is bounded because, by Lemma 5 and Eq. (C11),

$$\|\Pi X\|_\omega = \|\pi(X)\|_\tau \leqslant \|X\|_\omega. \quad \text{(C12)}$$

It is self-adjoint because, for any $X, Y \in L_2(\omega)$,

$$\langle Y, \Pi X \rangle_\omega = \langle Y, \pi(X) \otimes I \rangle_\omega = \langle \pi(Y), \pi(X) \rangle_\tau \quad \text{(C13)}$$

$$= \langle \pi(Y) \otimes I, X \rangle_\omega = \langle \Pi Y, X \rangle_\omega. \quad \text{(C14)}$$

It is also idempotent because, for any $X, Y \in L_2(\omega)$,

$$\langle Y, \Pi^2 X \rangle_\omega = \langle \Pi Y, \Pi X \rangle_\omega = \langle \pi(Y), \pi(X) \rangle_\tau \quad \text{(C15)}$$

$$= \langle Y, \Pi X \rangle_\omega, \quad \text{(C16)}$$

meaning that $\Pi^2 = \Pi$. Hence Π is a projection (see [25], Theorem 4.7.7) and Eq. (C9) is a basic property. ∎

Proof of Lemma 3. Let $\mathfrak{T}(\mathcal{H})$ be the Banach space of all trace-class self-adjoint operators on $\mathcal{H}$ [34]. The partial trace $\text{tr}' : \mathfrak{T}(\mathcal{H} \otimes \mathcal{H}') \to \mathfrak{T}(\mathcal{H})$ is a bounded linear map (see [30], p. 150) and therefore continuous (see [25], Theorem 1.5.7). It follows that tr' is differentiable and the derivative (as a linear map) is given by tr' itself (see [26], Example 8.1.3). As $\{\omega_\theta\}$ is assumed to be regular, ω_θ is differentiable at the truth.

Thus, $\tau_\theta = \mathrm{tr}' \omega_\theta$ is also differentiable at the truth (see [26], Theorem 8.2.1) and $\dot\tau \in \mathfrak{T}(\mathcal{H})$ is determined by the chain rule

$$\dot\tau = \mathrm{tr}' \dot\omega. \tag{C17}$$

Since $\{\omega_\theta\}$ is assumed to be regular, its score $S \in L_2(\omega)$ exists. The definition of the partial trace as per Eq. (C3) and the definition of the score as per Eq. (3.6) can then be used to give

$$\mathrm{tr}\, u\dot\tau = \mathrm{tr}(u \otimes I)\dot\omega = \langle u \otimes I, S \rangle_\omega \ \forall u \in \mathcal{B}(\mathcal{H}). \tag{C18}$$

The linear functional $\mathrm{tr}\, u\dot\tau = \langle u \otimes I, S \rangle_\omega$ of u is bounded with respect to $\|u\|_\tau$ by Eq. (C10) and can therefore be extended uniquely to be defined on $L_2(\tau)$. The score $\mathrm{tr}'_* S \in L_2(\tau)$ of $\{\tau_\theta\}$ then exists by definition, $\{\tau_\theta\}$ is regular, and

$$\langle u, \mathrm{tr}'_* S \rangle_\tau = \langle u \otimes I, S \rangle_\omega \ \forall u \in L_2(\tau). \tag{C19}$$

By Lemma 6, the score must be equal to the conditional expectation

$$\mathrm{tr}'_* S = \pi(S), \tag{C20}$$

which observes the shrinking property

$$\|\mathrm{tr}'_* S\|_\tau = \|\pi(S)\|_\tau \leqslant \|S\|_\omega. \tag{C21}$$

∎

Remark. Throughout this paper, differentiability is always assumed to be in the Fréchet sense [27].

APPENDIX D: PROOF OF THEOREM 1

Before proving Theorem 1, I need a few lemmas. The first lemma gives $\mathrm{supp} P_\theta = \mathrm{supp} P_0$ under general conditions.

Lemma 7. If $g_\theta(x) > 0$ a.e. P_0 on $\mathrm{supp} P_0$, the P_θ defined by Eq. (4.5) and P_0 have the same support.

Proof of Lemma 7. Write $C = \mathrm{supp} P_0$ for brevity in this proof. First notice that, since $P_0(C) = 1$,

$$P_\theta(C) = \int_C g_\theta(x) P_0(dx) = \int g_\theta(x) P_0(dx) = 1, \tag{D1}$$

and the first condition for $\mathrm{supp} P_\theta = C$ is satisfied. To prove the second condition, consider, for any closed strict subset D of C,

$$P_\theta(D) = \int_D g_\theta(x) P_0(dx) = 1 - \int_{C-D} g_\theta(x) P_0(dx). \tag{D2}$$

By the definition of C, $P_0(D) < 1$ and $P_0(C - D) > 0$. I can then prove that the last integral in Eq. (D2) is not zero by contradiction. Suppose that it is zero. Then, by Proposition 4.1.7 in Ref. [31],

$$g_\theta(x) = 0 \quad \text{a.e. } P_0 \text{ on } C - D. \tag{D3}$$

This statement contradicts the assumption that $g_\theta(x) > 0$ a.e. P_0 on C by the following argument. The assumption implies the existence of a set Y_1 with $P_0(Y_1) = 0$ such that

$$g_\theta(x) > 0 \ \forall x \in C - Y_1. \tag{D4}$$

On the other hand, Eq. (D3) implies the existence of a Y_2 with $P_0(Y_2) = 0$ such that

$$g_\theta(x) = 0 \ \forall x \in C - D - Y_2. \tag{D5}$$

The common domain $(C - Y_1) \cap (C - D - Y_2) = C - D - (Y_1 \cup Y_2)$ is not empty because $P_0[C - D - (Y_1 \cup Y_2)] = P_0(C - D) > 0$. Thus, for any x in the common domain, the two statements in Eqs. (D4) and (D5) contradict each other. By contradiction, the last integral in Eq. (D2) must be strictly positive, $P_\theta(D) < 1$, and the second condition for $\mathrm{supp} P_\theta = C$ is also satisfied. ∎

The next lemma provides some bounds on $g_\theta(x)$ given by Eqs. (4.6) and (4.14), which lead to useful properties for the submodels in Secs. IV and V.

Lemma 8. For the $g_\theta(x)$ given by Eqs. (4.6) and (4.14), there exist finite positive constants c_3, c_4, c_5 such that, for all $|\theta| \leqslant c < \infty$ and $|x| \leqslant \Delta < \infty$,

$$0 < c_3 \leqslant g_\theta(x) \leqslant c_4 < \infty, \tag{D6}$$

$$|\partial g_\theta(x)| \leqslant c_5 < \infty, \tag{D7}$$

where ∂ denotes $\partial/\partial\theta$. Moreover, defining

$$\bar u(\theta) \equiv \int u(x) g_\theta(x) P_0(dx) \tag{D8}$$

for a P_0 that obeys Eq. (2.17) and a P_0-integrable function $u(x)$, ∂ can be taken inside the integral, viz.,

$$\partial \bar u(\theta) = \int u(x) \partial g_\theta(x) P_0(dx). \tag{D9}$$

Proof of Lemma 8. Let

$$\gamma_\theta(x) \equiv 1 + \tanh[\theta a_\mu(x)] \tag{D10}$$

be the numerator of $g_\theta(x)$. Here $\gamma_\theta(x)$ is obviously continuous with respect to θ and x. For all finite θ and x,

$$0 < \gamma_\theta(x) < 2. \tag{D11}$$

By the extreme value theorem (see [24], Theorem 4.16), there exist $\theta_1, \theta_2 \in [-c, c]$ and $x_1, x_2 \in [-\Delta, \Delta]$ such that

$$\inf_{|\theta| \leqslant c, |x| \leqslant \Delta} \gamma_\theta(x) = \gamma_{\theta_1}(x_1), \tag{D12}$$

$$\sup_{|\theta| \leqslant c, |x| \leqslant \Delta} \gamma_\theta(x) = \gamma_{\theta_2}(x_2). \tag{D13}$$

It follows that

$$0 < c_3 \equiv \frac{\gamma_{\theta_1}(x_1)}{\gamma_{\theta_2}(x_2)} \leqslant g_\theta(x) \leqslant \frac{\gamma_{\theta_2}(x_2)}{\gamma_{\theta_1}(x_1)} \equiv c_4 < \infty. \tag{D14}$$

The proof of Eq. (D7) is similar, by noting that $\partial \gamma_\theta(x)$ is also continuous and bounded. Equation (D7) then implies Eq. (D9) by a corollary of the dominated convergence theorem (see [28], Corollary 2.8.7). ∎

Next I define what it means exactly for a quantity to be a function of Δ and also normalize the $H, A, L, \dot H$ matrices to remove their dependence on Δ, for later use.

Definition 1. Let a function $T_\Delta : \mathbb{R} \to \mathbb{R}$ be

$$T_\Delta(x) \equiv \frac{x}{\Delta}. \tag{D15}$$

Define a standard measure R_θ on $(\mathbb{R}, \Sigma)$ to describe the distribution of the object with standard size as

$$R_\theta(\cdot) = P_\theta \big[T_\Delta^{-1}(\cdot) \big] \tag{D16}$$

such that

$$\sup_{y \in \text{supp} R_\theta} |y| = 1. \tag{D17}$$

As T_Δ is invertible,

$$P_\theta(\cdot) = R_\theta[T_\Delta(\cdot)]. \tag{D18}$$

A function of the true P_0 is said to be a function of Δ if R_0 is fixed while P_0 varies with Δ through Eq. (D18).

The Hankel matrix for R_θ is

$$G_{pq}(\theta) \equiv \langle y^p, y^q \rangle_{R_\theta} = \Delta^{-p-q} H_{pq}(\theta). \tag{D19}$$

The orthonormal polynomials with respect to R_θ, satisfying $\langle b_n, b_m \rangle_{R_\theta} = \delta_{nm}$, are given by

$$b_n(y) = \sum_{p=0}^{n} B_{np}(\theta) y^p, \quad B_{np}(\theta) = A_{np}(\theta) \Delta^p. \tag{D20}$$

Let $V(\theta)$ be the Cholesky factor of $G(\theta)$. It is given by

$$V_{pn}(\theta) = \langle y^p, b_n \rangle_{R_\theta} = \Delta^{-p} L_{pn}(\theta). \tag{D21}$$

At $\theta = 0$, assuming the score given by Eq. (4.14), $\dot{G}_{pq}$ is given by

$$\dot{G}_{pq} = \langle y^{p+q}, b_\mu \rangle_{R_0} = V_{p+q\,\mu}(0). \tag{D22}$$

I need another lemma to evaluate the derivative of the purification given by Eq. (5.6).

Lemma 9. Assume the conditions for Theorem 1 and consider the submodel given by Eqs. (4.5)–(4.7), (4.14), (5.6), and (5.7). Let

$$|\Phi_\theta^j\rangle \equiv \sum_{p=0}^{j} \sum_{n=0}^{p} \frac{(-i\hat{k})^p}{p!} L_{pn}(\theta) |\psi\rangle \otimes |n\rangle \tag{D23}$$

be a function of $\theta \in (-c, c)$ with values in $\mathcal{H} \otimes \mathcal{H}''$.

(i) The $\{|\Phi_\theta^j\rangle\}$ converges uniformly to the $|\Phi_\theta\rangle$ given by Eq. (5.6) in the sense that

$$\lim_{j \to \infty} \sup_{\theta \in (-c,c)} \left\| |\Phi_\theta^j\rangle - |\Phi_\theta\rangle \right\| = 0. \tag{D24}$$

(ii) For any j and θ, the derivative of Eq. (D23) with respect to θ exists in $\mathcal{H} \otimes \mathcal{H}''$ and is given by

$$\partial |\Phi_\theta^j\rangle = \sum_{p=0}^{j} \sum_{n=0}^{p} \frac{(-i\hat{k})^p}{p!} \partial L_{pn}(\theta) |\psi\rangle \otimes |n\rangle. \tag{D25}$$

(iii) The $\partial |\Phi_\theta\rangle$ exists in $\mathcal{H} \otimes \mathcal{H}''$, and $\{\partial |\Phi_\theta^j\rangle\}$ converges uniformly to it.

Proof of Lemma 9. Since each measure in the Szegő class has an infinite and bounded support by definition, P_0 satisfies the condition for Lemma 1 by assumption. By Lemmas 7 and 8, $\text{supp} P_\theta = \text{supp} P_0$, so Lemma 1 can be applied to the whole submodel and Eqs. (5.6) and (D23) are well defined. To prove statement (ii) of the lemma, note that ∂L_{pn} is given by the formula (see [42], Theorem A.1)

$$\partial L_{pn} = \Delta^p \partial V_{pn}, \tag{D26}$$

$$\partial V_{pn} = \sum_{m=0}^{p} \sum_{q=0}^{m} \sum_{r=0}^{n} V_{pm} W_{mn} B_{mq}(\partial G_{qr}) B_{nr}, \tag{D27}$$

$$W_{mn} \equiv \begin{cases} 1, & n < m \\ \frac{1}{2}, & n = m \\ 0, & n > m, \end{cases} \tag{D28}$$

in terms of the normalized quantities in Definition 1. By Lemma 8 and Eq. (2.17),

$$|\partial G_{qr}| = \left| \int \left(\frac{x}{\Delta}\right)^{q+r} \partial g_\theta(x) P_0(dx) \right| \leqslant c_5, \tag{D29}$$

so $|\partial L_{pn}| < \infty$, $\partial |\Phi_\theta^j\rangle$ exists, and the basic rules of differentiation give Eq. (D25) (see [27], Theorems 1.1.3 and 1.1.4).

To prove statements (i) and (iii), the plan is to prove that the sequence of functions satisfy the two conditions for Theorem 1.7.1 in Ref. [27], which implies the statements. The first condition is that $\{|\Phi_\theta^j\rangle\}$ converges to $|\Phi_\theta\rangle$ at a point θ. This condition is satisfied, as Lemmas 7, 8, and 1 imply the pointwise convergence at any θ. The second condition is the uniform convergence of $\{\partial |\Phi_\theta^j\rangle\}$, which I prove next. Suppose $J > j$ and consider

$$\| \partial |\Phi_\theta^J\rangle - \partial |\Phi_\theta^j\rangle \|^2$$

$$= \left\| \sum_{p=j+1}^{J} \sum_{n=0}^{J} \frac{(-i\hat{k})^p}{p!} \partial L_{pn} |\psi\rangle \otimes |n\rangle \right\|^2 \tag{D30}$$

$$= \sum_{n=0}^{J} \sum_{q=j+1}^{J} \sum_{p=j+1}^{J} \frac{i^q(-i)^p}{q!p!} \langle \psi | \hat{k}^{q+p} |\psi\rangle (\partial L_{qn})(\partial L_{pn}) \tag{D31}$$

$$\leqslant \sum_{n=0}^{J} \left(\sum_{p=j+1}^{J} \frac{\|\hat{k}\|_{\text{op}}^p}{p!} \partial L_{pn} \right)^2 \tag{D32}$$

$$\leqslant \sum_{n=0}^{J} \left[\sum_{p=j+1}^{J} \frac{(\|\hat{k}\|_{\text{op}} \Delta)^{2p}}{p!} \right] \sum_{p=j+1}^{J} \frac{(\partial V_{pn})^2}{p!} \tag{D33}$$

$$= \underbrace{\left[\sum_{p=j+1}^{J} \frac{(\|\hat{k}\|_{\text{op}} \Delta)^{2p}}{p!} \right]}_{s_1^J - s_1^j} \underbrace{\sum_{p=j+1}^{J} \frac{1}{p!} \sum_{n=0}^{p} (\partial V_{pn})^2}_{s_2^J(\theta) - s_2^j(\theta)}, \tag{D34}$$

where the right-hand side of Eq. (D30) has used the lower triangularity of ∂L_{pn} to replace $\sum_{n=0}^{p}$ by $\sum_{n=0}^{J}$, Eq. (D32) has used the series of absolute values to bound Eq. (D31) and the definition of the operator norm to bound $|\langle \psi | \hat{k}^{q+p} |\psi\rangle| \leqslant \|\hat{k}\|_{\text{op}}^{q+p}$, Eq. (D33) has used the CSI and Eq. (D26), and s_1^j and $s_2^j(\theta)$ are defined as

$$s_1^j \equiv \sum_{p=0}^{j} \frac{(\|\hat{k}\|_{\text{op}} \Delta)^{2p}}{p!}, \tag{D35}$$

$$s_2^j(\theta) \equiv \sum_{p=0}^{j} \frac{1}{p!} \sum_{n=0}^{p} [\partial V_{pn}(\theta)]^2. \tag{D36}$$

If $\{s_1^j\}$ converges and $\{s_2^j(\theta)\}$ converges uniformly, then the left-hand side of Eq. (D30) can be bounded uniformly. $\{s_1^j\}$ converges to $\exp[(\|\hat{k}\|_{\text{op}} \Delta)^2]$ for a bounded $\hat{k}$, so it remains to be proved that $\{s_2^j(\theta)\}$ converges uniformly.

To bound $(\partial V_{pn})^2$, apply the CSI to Eq. (D27) to obtain

$$(\partial V_{pn})^2 \leqslant \left[\sum_{m=0}^{p} (V_{pm})^2 \right] \sum_{m=0}^{p} (W_{mn} D_{mn})^2, \qquad \text{(D37)}$$

$$D \equiv B_{(p)} [\partial G_{(p)}] B_{(p)}^\top, \qquad \text{(D38)}$$

$$\sum_{n=0}^{p} (\partial V_{pn})^2 \leqslant G_{pp} \sum_{n=0}^{p} \sum_{m=0}^{p} (W_{mn} D_{mn})^2 \qquad \text{(D39)}$$

$$\leqslant \frac{G_{pp}}{2} \|D\|_{\mathrm{HS}}^2 \leqslant \frac{1}{2} \|D\|_{\mathrm{HS}}^2, \qquad \text{(D40)}$$

where $G_{pp} = \sum_{m=0}^{p} (V_{pm})^2$ in Eq. (D39) comes from the fact that V is the Cholesky factor of G, $\|\cdot\|_{\mathrm{HS}}$ is the Hilbert-Schmidt norm (also called the Frobenius norm)

$$\|D\|_{\mathrm{HS}}^2 \equiv \sum_{m,n} (D_{mn})^2, \qquad \text{(D41)}$$

the first bound in Eq. (D40) comes from the definition of W and the symmetry of D [see [12], Eq. (B34)], and the last bound in Eq. (D40) comes from

$$G_{pp} = \|y^p\|_{R_\theta}^2 \leqslant 1, \qquad \text{(D42)}$$

by Eq. (D17). $\|D\|_{\mathrm{HS}}$ can be further bounded in terms of $\|\partial G_{(p)}\|_{\mathrm{HS}}$ and the operator norm $\|B_{(p)}\|_{\mathrm{op}}$ as (see [64], Theorems II.2.11 and II.3.9)

$$\|D\|_{\mathrm{HS}} \leqslant \|\partial G_{(p)}\|_{\mathrm{HS}} \|B_{(p)}\|_{\mathrm{op}}^2 = \frac{\|\partial G_{(p)}\|_{\mathrm{HS}}}{\lambda_{\min}[G_{(p)}]}, \qquad \text{(D43)}$$

where $\lambda_{\min}$ denotes the smallest eigenvalue and the last step has used $B_{(p)}^\top B_{(p)} = G_{(p)}^{-1}$ from Lemma 4 and Eq. (II.2.10) in Ref. [64] to obtain

$$\|B_{(p)}\|_{\mathrm{op}}^2 = \lambda_{\max}[G_{(p)}^{-1}] = \frac{1}{\lambda_{\min}[G_{(p)}]}. \qquad \text{(D44)}$$

To bound $\lambda_{\min}[G_{(p)}]$, first note that, for any column vector $u \in \mathbb{R}^p$,

$$u^\top G_{(p)}(\theta) u = \int \left[\sum_{q=0}^{p} u_q \left(\frac{x}{\Delta} \right)^q \right]^2 g_\theta(x) P_0(dx) \qquad \text{(D45)}$$

$$\geqslant c_3 u^\top G_{(p)}(0) u, \qquad \text{(D46)}$$

where $c_3 > 0$ comes from Lemma 8. Applying the Rayleigh quotient theorem [65]

$$\lambda_{\min}[G_{(p)}(\theta)] = \min_{u^\top u = 1} u^\top G_{(p)}(\theta) u \qquad \text{(D47)}$$

to Eq. (D46), I obtain

$$\lambda_{\min}[G_{(p)}(\theta)] \geqslant c_3 \lambda_{\min}[G_{(p)}(0)]. \qquad \text{(D48)}$$

If P_0 is in the Szegő class, R_0 is also in the Szegő class, and by a theorem of Widom and Wilf [51] there exists a constant $0 < r < 1$ such that $\lambda_{\min}$ of its Hankel matrix satisfies

$$\lambda_{\min}[G_{(p)}(0)] = \Theta(\sqrt{p} r^p), \qquad \text{(D49)}$$

where the Θ order here is in terms of $p \to \infty$. To bound the other quantity $\|\partial G_{(p)}\|_{\mathrm{HS}}$ in Eq. (D43), use Eq. (D29) to write

$$\|\partial G_{(p)}\|_{\mathrm{HS}}^2 \leqslant \sum_{n=0}^{p} \sum_{m=0}^{p} c_5^2 = c_5^2 (p+1)^2. \qquad \text{(D50)}$$

Putting everything together, $s_2^j(\theta)$ in Eq. (D36) can be bounded as

$$s_2^j(\theta) \leqslant \sum_{p=0}^{j} \frac{c_5^2 (p+1)^2}{p! 2 [c_3 \Theta(\sqrt{p} r^p)]^2}. \qquad \text{(D51)}$$

The right-hand side does not depend on θ and converges by the ratio test, so $\{s_2^j(\theta)\}$ converges uniformly by the Weierstrass test (see [24], Theorem 7.10).

With the convergence of $\{s_1^j\}$ and the uniform convergence of $\{s_2^j(\theta)\}$, given any $\varepsilon > 0$, there exists a θ-independent j_0 such that, for all $\theta \in (-c, c)$ and $J \geqslant j \geqslant j_0$, $s_1^J - s_1^j < \varepsilon$, $s_2^J(\theta) - s_2^j(\theta) < \varepsilon$, and by Eqs. (D30)–(D34),

$$\left\| \partial \left| \Phi_\theta^J \right\rangle - \partial \left| \Phi_\theta^j \right\rangle \right\| < \varepsilon, \qquad \text{(D52)}$$

$$\sup_{\theta \in (-c, c)} \left\| \partial \left| \Phi_\theta^J \right\rangle - \partial \left| \Phi_\theta^j \right\rangle \right\| \leqslant \varepsilon, \qquad \text{(D53)}$$

which means that $\{\partial |\Phi_\theta^j\rangle\}$ is Cauchy in the space of bounded Hilbert-space-valued functions $\mathcal{B}_{\mathcal{H} \otimes \mathcal{H}''}[(-c, c)]$ with the supremum norm $\sup_{\theta \in (-c, c)} \| |u_\theta\rangle \|$ for each $|u_\theta\rangle \in \mathcal{B}_{\mathcal{H} \otimes \mathcal{H}''}[(-c, c)]$ (see [26], Sec. 7.1). As the Hilbert space $\mathcal{H} \otimes \mathcal{H}''$ is a Banach space, $\mathcal{B}_{\mathcal{H} \otimes \mathcal{H}''}[(-c, c)]$ is also a Banach space (see [26], Theorem 7.1.3), the completeness of which means that the Cauchy $\{\partial |\Phi_\theta^j\rangle\}$ converges uniformly. Hence, the two conditions for Theorem 1.7.1 in Ref. [27] are satisfied and the theorem implies statements (i) and (iii) of the lemma here. ∎

Proof of Theorem 1. Consider the submodel given by Eqs. (4.5)–(4.7), (4.14), (5.6), and (5.7). Let $|\dot{\Phi}\rangle \equiv \partial |\Phi_\theta\rangle|_{\theta=0}$. The Helstrom information of the purified model is the Fubini-Study metric [66]

$$\left\| S_\Phi \right\|_{\Phi_0}^2 = 4 \left(\langle \dot{\Phi} | \dot{\Phi} \rangle - |\langle \dot{\Phi} | \Phi_0 \rangle|^2 \right) \leqslant 4 \langle \dot{\Phi} | \dot{\Phi} \rangle. \qquad \text{(D54)}$$

Lemma 9 implies that, under the conditions for Theorem 1, $\langle \dot{\Phi} | \dot{\Phi} \rangle < \infty$, $\{ |\dot{\Phi}^j\rangle \equiv \partial |\Phi_\theta^j\rangle|_{\theta=0} \}$ converges to $|\dot{\Phi}\rangle$, and therefore the squared norm also converges to [see [25], Eq. (3.9)]

$$\langle \dot{\Phi}^j | \dot{\Phi}^j \rangle \to \langle \dot{\Phi} | \dot{\Phi} \rangle < \infty. \qquad \text{(D55)}$$

With Eqs. (2.7), (2.8), and (D25), I can write

$$\langle \dot{\Phi}^j | \dot{\Phi}^j \rangle = \int \sum_{n=0}^{j} \left| \sum_{p=0}^{j} \frac{(-ik)^p}{p!} \dot{L}_{pn} \right|^2 Q(dk). \qquad \text{(D56)}$$

To evaluate this expression as a function of Δ, write the formula for $\dot{L}_{pn}$ in Eqs. (D26) and (D27) at $\theta = 0$ as

$$\dot{L}_{pn} = \Delta^p \sum_{m=0}^{p} \sum_{q=0}^{m} \sum_{r=0}^{n} V_{pm} W_{mn} B_{mq} V_{q+r\,\mu} B_{nr}, \qquad \text{(D57)}$$

where $\dot{G}_{qr} = V_{q+r\,\mu}$ comes from Eq. (D22). As V, W, B are all lower triangular, for $\dot{L}_{pn}$ to be nonzero, the indices in Eq. (D57) should satisfy

$$p \geqslant m \geqslant n \geqslant r, \quad m \geqslant q, \quad q + r \geqslant \mu, \qquad \text{(D58)}$$

which imply $2m \geqslant q + m \geqslant q + r \geqslant \mu$ and

$$p \geqslant m \geqslant \left\lceil \frac{\mu}{2} \right\rceil. \qquad \text{(D59)}$$

Thus,

$$\dot{L}_{pn} = 0 \quad \text{if } p < \left\lceil \frac{\mu}{2} \right\rceil. \tag{D60}$$

With $\dot{L}_{pn} \propto \Delta^p$, Eq. (D56) is a power series of Δ. By Eq. (D55), the power series converges, so I can conclude from Eqs. (D55), (D56), and (D60) that [50]

$$\langle \dot{\Phi} | \dot{\Phi} \rangle = O(\Delta^{2\lceil \mu/2 \rceil}). \tag{D61}$$

To evaluate the Helstrom bound, I also need $\dot{\beta}$. Given Eqs. (4.12), (4.14), and (5.10), it is given by

$$\dot{\beta} = \langle a_\mu, x^\mu + o(\Delta^\mu) \rangle_{P_0} = \Delta^\mu V_{\mu\mu} + \langle a_\mu, o(\Delta^\mu) \rangle_{P_0}$$
$$= \Theta(\Delta^\mu), \tag{D62}$$

where the second equality comes from Lemma 4 and Eq. (D21) and the final equality comes from $V_{\mu\mu} > 0$ by Lemma 4 and the CSI

$$|\langle a_\mu, o(\Delta^\mu) \rangle_{P_0}| \leqslant \|a_\mu\|_{P_0} \|o(\Delta^\mu)\|_{P_0} = o(\Delta^\mu). \tag{D63}$$

Finally, the theorem is given by

$$\tilde{\mathsf{H}} \geqslant \mathsf{H} \geqslant \frac{\dot{\beta}^2}{N \|S_\Phi\|_{\Phi_0}^2} \geqslant \frac{\dot{\beta}^2}{4N \langle \dot{\Phi} | \dot{\Phi} \rangle} = \frac{\Omega(\Delta^{2\lfloor \mu/2 \rfloor})}{N}, \tag{D64}$$

where the first inequality comes from the submodel bound in Lemma 2, the second inequality comes from applying the purification bound in Lemma 3 to the purification in Lemma 1, the third inequality comes from Eq. (D54), and the final equality comes from Eqs. (D61) and (D62). ∎

Remark. Although the assumption of an infinite and bounded suppP_0 is central to Lemma 1, Lemma 8, and Theorem 1, the more specific Szegő-class assumption is used only in Eqs. (D49) and (D51) in Lemma 9, to ensure the convergence of the Helstrom information of the purified model. If $\lambda_{\min}[G_{(p)}(0)]$ can be lower bounded in another way that still ensures the uniform convergence of $\{s_2^j(\theta)\}$ in Eq. (D51), then Lemma 9 and Theorem 1 still hold and the Szegő-class assumption may be relaxed. For example, Theorem 3 in Ref. [67] gives the asymptotic behavior of $\lambda_{\min}[G_{(p)}(0)]$ for a more general class of P_0 that still makes $\{s_2^j(\theta)\}$ converge uniformly, although the conditions there are much harder to state or check.

APPENDIX E: PROOF OF PROPOSITION 2

Proof of Proposition 2. Let $\nu[P]$ be the probability measure for each photon, with a probability density given by Eq. (7.1), and let $\{\nu[P_\theta] : \theta \in (-c, c)\}$ be the 1D submodel based on the $\{P_\theta\}$ given by Eqs. (4.5)–(4.7) and (4.14). The score $\nu_* S$ of $\{\nu[P_\theta]\}$ is given by

$$(\nu_* S)(\xi) = \frac{\dot{\eta}(\xi)}{\eta_0(\xi)}, \tag{E1}$$

$$\eta_\theta(\xi) \equiv \eta(\xi, P_\theta) = \int h(\xi - x) P_\theta(dx), \tag{E2}$$

$$\dot{\eta}(\xi) = \int h(\xi - x) a_\mu(x) P_0(dx). \tag{E3}$$

Assume the normalized quantities defined in Definition 1 in the following. Using Taylor's theorem [24], $h(\xi - x) =$

$h(\xi - \Delta y)$ becomes

$$h(\xi - \Delta y) = \sum_{p=0}^{\mu} \frac{h^{(p)}(\xi)}{p!} (-\Delta y)^p + r(\xi, y) \Delta^{\mu+1}, \tag{E4}$$

where the remainder is given by

$$r(\xi, y) \equiv \frac{h^{(\mu+1)}(\xi - \Delta \bar{y})}{(\mu + 1)!} (-y)^{\mu+1} \tag{E5}$$

for a certain $\bar{y}(\xi, y)$ between 0 and y. Equation (E3) becomes

$$\dot{\eta}(\xi) = \frac{h^{(\mu)}(\xi)}{\mu!} (-\Delta)^\mu V_{\mu\mu} + \Delta^{\mu+1} \langle r(\xi, y), b_\mu \rangle_R. \tag{E6}$$

The last term can be bounded as

$$|\langle r(\xi, y), b_\mu \rangle_R| \leqslant \|r(\xi, y)\|_R \tag{E7}$$

$$\leqslant \frac{\bar{h}(\xi)}{(\mu + 1)!} \|y^{\mu+1}\|_R \tag{E8}$$

$$\leqslant \frac{\bar{h}(\xi)}{(\mu + 1)!}, \tag{E9}$$

where Eq. (E7) has used the CSI, Eq. (E8) has used Eq. (7.5) to bound the $|h^{(\mu+1)}(\xi - \Delta \bar{y})|$ in Eq. (E5), and Eq. (E9) comes from Eq. (D17). Apply the triangle inequality to Eq. (E6) to obtain

$$|\dot{\eta}(\xi)| \leqslant \left| \frac{h^{(\mu)}(\xi)}{\mu!} (-\Delta)^\mu V_{\mu\mu} \right| + |\Delta^{\mu+1} \langle r(\xi, y), b_\mu \rangle_R| \tag{E10}$$

$$\leqslant \frac{\Delta^\mu}{\mu!} \left[V_{\mu\mu} |h^{(\mu)}(\xi)| + \frac{\Delta \bar{h}(\xi)}{\mu + 1} \right]. \tag{E11}$$

The norm of the score given by Eq. (E1) can be bounded again by the triangle inequality as

$$\left\| \frac{\dot{\eta}}{\eta_0} \right\|_{\nu[P_0]} \leqslant \frac{\Delta^\mu}{\mu!} \left[V_{\mu\mu} \left\| \frac{h^{(\mu)}}{\eta_0} \right\|_{\nu[P_0]} + \frac{\Delta}{\mu + 1} \left\| \frac{\bar{h}}{\eta_0} \right\|_{\nu[P_0]} \right]. \tag{E12}$$

Equation (7.4) implies that

$$\eta_0(\xi) \geqslant \underline{h}(\xi), \tag{E13}$$

while Eqs. (7.6) imply that

$$\left\| \frac{h^{(\mu)}}{\eta_0} \right\|_{\nu[P_0]}^2 \leqslant \int \frac{[h^{(\mu)}(\xi)]^2}{\underline{h}(\xi)} d\xi < \infty, \tag{E14}$$

$$\left\| \frac{\bar{h}}{\eta_0} \right\|_{\nu[P_0]}^2 \leqslant \int \frac{[\bar{h}(\xi)]^2}{\underline{h}(\xi)} d\xi < \infty. \tag{E15}$$

The convergence of these quantities means that Eq. (E12) can be written as

$$\left\| \frac{\dot{\eta}}{\eta_0} \right\|_{\nu[P_0]} = O(\Delta^\mu) \tag{E16}$$

and the Fisher information of the submodel becomes

$$\|\nu_* S\|_{\nu[P_0]}^2 = \left\| \frac{\dot{\eta}}{\eta_0} \right\|_{\nu[P_0]}^2 = O(\Delta^{2\mu}). \tag{E17}$$

Given this expression and Eq. (D62), the Cramér-Rao bound for the M-temporal-mode submodel $\{\nu^{(M,\epsilon)}[P_\theta] :$

$\theta \in (-c, c)\}$ is

$$\mathsf{C} = \frac{\dot{\beta}^2}{N \| \nu_* S \|^2_{\nu[P_0]}} = \frac{\Omega(1)}{N}, \qquad (E18)$$

and the fact that any submodel bound is a lower bound on the semiparametric bound $\tilde{\mathsf{C}}$ (see [43], Lemma 25.19) leads to Eq. (7.7). ∎

[1] G. de Villiers and E. R. Pike, *The Limits of Resolution* (CRC Press, Boca Raton, 2016).

[2] J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).

[3] *Handbook of Biological Confocal Microscopy*, edited by J. B. Pawley (Springer, New York, 2006).

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

[5] M. Tsang, Resolving starlight: A quantum perspective, Contemp. Phys. **60**, 279 (2019).

[6] M. Tsang, R. Nair, and X.-M. Lu, Quantum Theory of Superresolution for Two Incoherent Optical Point Sources, Phys. Rev. X **6**, 031033 (2016).

[7] F. Yang, A. Tashchilina, E. S. Moiseev, C. Simon, and A. I. Lvovsky, Far-field linear optical superresolution via heterodyne detection in a higher-order local oscillator mode, Optica **3**, 1148 (2016).

[8] M. Tsang, Subdiffraction incoherent optical imaging via spatial-mode demultiplexing, New J. Phys. **19**, 023054 (2017).

[9] M. Tsang, Subdiffraction incoherent optical imaging via spatial-mode demultiplexing: Semiclassical treatment, Phys. Rev. A **97**, 023830 (2018).

[10] M. Tsang, Semiparametric estimation for incoherent optical imaging, Phys. Rev. Research **1**, 033006 (2019).

[11] S. Zhou and L. Jiang, Modern description of Rayleigh's criterion, Phys. Rev. A **99**, 013808 (2019).

[12] M. Tsang, Quantum limit to subdiffraction incoherent optical imaging, Phys. Rev. A **99**, 012305 (2019).

[13] K. A. G. Bonsma-Fisher, W.-K. Tham, H. Ferretti, and A. M. Steinberg, Realistic sub-Rayleigh imaging with phase-sensitive measurements, New J. Phys. **21**, 093010 (2019).

[14] K. Liang, S. A. Wadood, and A. N. Vamivakas, Coherence effects on estimating general sub-Rayleigh object distribution moments, Phys. Rev. A **104**, 022220 (2021).

[15] Z. Dutton, R. Kerviche, A. Ashok, and S. Guha, Attaining the quantum limit of superresolution in imaging an object's length via predetection spatial-mode sorting, Phys. Rev. A **99**, 033847 (2019).

[16] S. Prasad, Quantum limited superresolution of extended sources in one and two dimensions, Phys. Rev. A **102**, 063719 (2020).

[17] E. Bisketzi, D. Branford, and A. Datta, Quantum limits of localisation microscopy, New J. Phys. **21**, 123032 (2019).

[18] L. J. Fiderer, T. Tufarelli, S. Piano, and G. Adesso, General expressions for the quantum Fisher information matrix with applications to discrete quantum imaging, PRX Quantum **2**, 020308 (2021).

[19] K. K. M. Bearne, Y. Zhou, B. Braverman, J. Yang, S. A. Wadood, A. N. Jordan, A. N. Vamivakas, Z. Shi, and R. W. Boyd, Confocal super-resolution microscopy based on a spatial mode sorter, Opt. Express **29**, 11784 (2021).

[20] F. Bao, H. Choi, V. Aggarwal, and Z. Jacob, Quantum-accelerated imaging of N stars, Opt. Lett. **46**, 3045 (2021).

[21] A. A. Pushkina, G. Maltese, J. I. Costa-Filho, P. Patel, and A. I. Lvovsky, Super-resolution linear optical imaging in the far field, arXiv:2105.01743.

[22] E. F. Matlin and L. J. Zipp, Adaptive imaging of arbitrary thermal source distributions with near quantum-limited resolution, arXiv:2106.13332.

[23] M. R. Grace and S. Guha, Quantum-optimal object discrimination in sub-diffraction incoherent imaging, arXiv:2107.00673.

[24] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. (McGraw-Hill, New York, 1976).

[25] L. Debnath and P. Mikusiński, *Introduction to Hilbert Spaces with Applications* (Elsevier, Amsterdam, 2005).

[26] J. Dieudonné, *Foundations of Modern Analysis* (Academic, New York, 1969).

[27] T. M. Flett, *Differential Analysis: Differentiation, Differential Equations and Differential Inequalities* (Cambridge University Press, Cambridge, 1980).

[28] V. Bogachev, *Measure Theory* (Springer, Berlin, 2007).

[29] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I: Functional Analysis* (Academic, San Diego, 1980).

[30] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).

[31] K. R. Parthasarathy, *Introduction to Probability and Measure* (Hindustan Book Agency, New Delhi, 2005).

[32] K. R. Parthasarathy, *Probability Measures on Metric Spaces* (Academic, New York, 1967).

[33] K. R. Parthasarathy, *An Introduction to Quantum Stochastic Calculus* (Birkhäuser, Basel, 1992).

[34] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, Pisa, 2011).

[35] A. S. Holevo, *Statistical Structure of Quantum Theory* (Springer, Berlin, 2001).

[36] M. Tsang, F. Albarelli, and A. Datta, Quantum Semiparametric Estimation, Phys. Rev. X **10**, 031023 (2020).

[37] C. Fabre and N. Treps, Modes and states in quantum optics, Rev. Mod. Phys. **92**, 035005 (2020).

[38] P. Boucher, C. Fabre, G. Labroille, and N. Treps, Spatial optical mode demultiplexing as a practical tool for optimal transverse distance estimation, Optica **7**, 1621 (2020).

[39] M. Tsang, Poisson quantum information, Quantum **5**, 527 (2021).

[40] M. Hayashi, *Quantum Information Theory: Mathematical Foundation*, 2nd ed. (Springer, Berlin, 2017).

[41] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kołodyński, in *Progress in Optics*, edited by E. Wolf (Elsevier, Amsterdam, 2015) Vol. 60, Chap. 4, pp. 345–435.

[42] S. Särkkä, *Bayesian Filtering and Smoothing* (Cambridge University Press, Cambridge, 2013), Algorithm A.1.

[43] A. W. van der Vaart, *Asymptotic Statistics* (Cambridge University Press, Cambridge, 1998).

[44] J. A. Gross and C. M. Caves, One from many: Estimating a function of many parameters, J. Phys. A: Math. Theor. **54**, 014001 (2020).

[45] P. J. Bickel, C. A. J. Klaassen, Y. Ritov, and J. A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models* (Springer, New York, 1993).

[46] J. M. Lee, *Introduction to Smooth Manifolds* (Springer, New York, 2003).

[47] G. H. Golub and G. Meurant, *Matrices, Moments and Quadrature with Applications* (Princeton University Press, Princeton, 2010).

[48] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology, Nat. Phys. **7**, 406 (2011).

[49] D. E. Knuth, Big Omicron and big Omega and big Theta, ACM SIGACT News **8**, 18 (1976).

[50] P. D. Miller, *Applied Asymptotic Analysis* (American Mathematical Society, Providence, 2006).

[51] H. Widom and H. Wilf, Small eigenvalues of large Hankel matrices, Proc. Am. Math. Soc. **17**, 338 (1966).

[52] J. Řeháček, M. Paúr, B. Stoklasa, Z. Hradil, and L. L. Sánchez-Soto, Optimal measurements for resolution beyond the Rayleigh limit, Opt. Lett. **42**, 231 (2017).

[53] M. Falk, J. Hüsler, and R.-D. Reiss, *Laws of Small Numbers: Extremes and Rare Events*, 3rd ed. (Birkhäuser, Basel, 2011).

[54] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. (Springer, New York, 2005).

[55] M. Paúr, B. Stoklasa, J. Grover, A. Krzic, L. L. Sánchez-Soto, Z. Hradil, and J. Řeháček, Tempering Rayleigh's curse with PSF shaping, Optica **5**, 1177 (2018).

[56] M. Paúr, B. Stoklasa, D. Koutný, J. Řeháček, Z. Hradil, J. Grover, A. Krzic, and L. L. Sánchez-Soto, Reading out Fisher information from the zeros of the point spread function, Opt. Lett. **44**, 3114 (2019).

[57] R. D. Gill and B. Y. Levit, Applications of the van Trees inequality: A Bayesian Cramér-Rao bound, Bernoulli **1**, 59 (1995).

[58] M. Tsang, Conservative classical and quantum resolution limits for incoherent imaging, J. Mod. Opt. **65**, 1385 (2018).

[59] M. Tsang, Physics-inspired forms of the Bayesian Cramér-Rao bound, Phys. Rev. A **102**, 062217 (2020).

[60] M. D. Vidrighin, G. Donati, M. G. Genoni, X.-M. Jin, W. S. Kolthammer, M. S. Kim, A. Datta, M. Barbieri, and I. A. Walmsley, Joint estimation of phase and phase diffusion for quantum metrology, Nat. Commun. **5**, 3532 (2014).

[61] S. Ng, S. Z. Ang, T. A. Wheatley, H. Yonezawa, A. Furusawa, E. H. Huntington, and M. Tsang, Spectrum analysis with quantum dynamical systems, Phys. Rev. A **93**, 042121 (2016).

[62] K. Ohki, in *Building Foundations for Quantum Statistical Modeling*, edited by F. Tanaka and T. Sugiyama (Kyoto University, Kyoto, 2017), Vol. 2018, pp. 18–44.

[63] M. Tsang, Quantum analogs of the conditional expectation for retrodiction and smoothing: A unified view, arXiv:1912.02711.

[64] G. W. Stewart and J.-g. Sun, *Matrix Perturbation Theory* (Academic, Boston, 1990).

[65] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985), Theorem 4.2.2.

[66] A. Fujiwara and H. Nagaoka, Quantum Fisher metric and estimation for pure state models, Phys. Lett. A **201**, 119 (1995).

[67] H. Widom, Polynomials associated with measures in the complex plane, J. Math. Mech. **16**, 997 (1967).