

## Semiparametric estimation for incoherent optical imaging

Mankei Tsang <sup>\*</sup>*Department of Electrical and Computer Engineering, National University of Singapore,**4 Engineering Drive 3, Singapore 117583, Singapore**and Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117551, Singapore*

(Received 18 June 2019; published 2 October 2019)

The theory of semiparametric estimation offers an elegant way of computing the Cramér-Rao bound for a parameter of interest in the midst of infinitely many nuisance parameters. Here I apply the theory to the problem of incoherent imaging under the effects of diffraction and photon shot noise, where the object may consist of an arbitrary number of point sources and the moments of their distribution are the parameters of interest. Using a Hilbert-space formalism designed for Poisson processes, I derive exact semiparametric Cramér-Rao bounds and efficient estimators for both direct imaging and a quantum-inspired measurement method called spatial-mode demultiplexing (SPADE). The results establish the superiority of SPADE even when little prior information about the object is available.

DOI: [10.1103/PhysRevResearch.1.033006](https://doi.org/10.1103/PhysRevResearch.1.033006)

## I. INTRODUCTION

Two fundamental problems confront incoherent optical imaging: the diffraction limit [1,2] and the photon shot noise [3,4]. To quantify their effects on the resolution rigorously, the Cramér-Rao bound (CRB) on the error of parameter estimation [5] has been widely used, especially in astronomy and fluorescence microscopy [6–17]. Most previous studies, however, assume that the object has a simple specific shape, such as a point source or two, and only one or a few parameters of the object are unknown. Such parametric models may not be justifiable when there is little prior information about the object. Without a limit on the number of point sources in the object or a parametric model, the CRB seems intractable—an infinite number of parameters are needed to specify the object distribution, leading to a Fisher information matrix with infinitely many entries, and then the infinite-dimensional matrix has to be inverted to give the CRB. While there also exist many studies on superresolution that can deal with more general objects [17–20], they either ignore noise or use noise models that are too simplistic to capture the signal-dependent nature of photon shot noise.

To compute the CRB and to evaluate the efficiency of estimators for general objects, here I propose a theory of semiparametric estimation for incoherent optical imaging. Semiparametric estimation refers to the estimation of a parameter of interest in the presence of infinitely many other unknown “nuisance” parameters [21,22]. The method has found many applications in econometrics, biostatistics, and astrostatistics

[21]. A typical example is the estimation of the mean of a random variable when its probability density is assumed to have finite variance but otherwise is arbitrary. Thanks to a beautiful Hilbert-space formalism [21,22], the semiparametric theory is able to compute the CRB for such problems despite the infinite dimensionality and also evaluate the existence and efficiency of semiparametric estimators. Such problems are exactly the type that bedevil the study of imaging thus far, and here I show how the semiparametric theory can be used to yield similarly elegant results for optical imaging.

The optics problem of interest here is the far-field imaging of an object emitting spatially incoherent light [2,4], with the most important applications being optical astronomy [6–9] and fluorescence microscopy [10–14]. With a finite numerical aperture, the imaging system introduces a spatial bandwidth limit to the waves, otherwise known as the diffraction limit [1,2]. The standard measurement, called direct imaging, records the intensity of the light on the image plane. Recently, quantum information theory inspired the invention of an alternative measurement called spatial-mode demultiplexing (SPADE) [23], which has been shown theoretically [23–53] and experimentally [54–60] to be superior to direct imaging in resolving two sub-Rayleigh sources and estimating the size and moments of a subdiffraction object. Most of the aforementioned studies, however, assume parametric models for the object. Exceptions include Refs. [24–27,51,57], which consider the estimation of the moments of an arbitrary object, but the results there are not conclusive—only the CRB for direct imaging was computed exactly [25], while the CRB for SPADE was evaluated only approximately [24,25,27]. Another problem is the existence and efficiency of unbiased moment estimators; again, only approximate results have been obtained so far [24,25]. Building on the established semiparametric theory [21,22], here I compute the exact semiparametric CRBs and also propose unbiased and efficient moment estimators for both direct imaging and SPADE. These results enable a fair and rigorous comparison of the two measurement

<sup>\*</sup>mankei@nus.edu.sg; <https://blog.nus.edu.sg/mankei/>

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

methods, which proves the fundamental superiority of SPADE for moment estimation, even when the object shape is unknown.

This paper is organized as follows. Section II introduces the Fisher information and the CRB for Poisson processes. Section III presents the semiparametric CRB in terms of a Hilbert-space formalism designed for such processes. Section IV introduces the models of direct imaging and SPADE. Section V computes the CRB for moment estimation with direct imaging and proposes an efficient estimator. Section VI shows how the CRB should be modified for a normalized object distribution. Section VII computes the CRB for SPADE and also proposes an efficient estimator. Section VIII uses the CRBs to compare the performances of direct imaging and SPADE, demonstrating the superiority of SPADE for subdiffraction objects. Section IX concludes the paper and points out open issues, while the Appendixes detail the technical issues that arise in the main text.

## II. CRAMÉR-RAO BOUND FOR POISSON PROCESSES

For optical astronomy [4,6,8,9], fluorescence microscopy [10–14], and even electron microscopy [15,16], Poisson noise can be safely assumed. Suppose that each detector in a photodetector array is labeled by  $x \in \mathcal{X}$ , where  $\mathcal{X}$  denotes the detector space. Assume that the observed process, such as the image recorded by a camera, is a Poisson random measure  $n$  on  $\mathcal{X}$  and its  $\sigma$ -algebra  $\Sigma$ , with a mean given by the intensity measure  $\bar{n}$  on the same  $(\mathcal{X}, \Sigma)$  [61].  $n(\mathcal{A})$  for any  $\mathcal{A} \in \Sigma$  is then a Poisson variable with mean  $\bar{n}(\mathcal{A})$ . For example, if  $\mathcal{X} \subseteq \mathbb{R}^2$  is a two-dimensional surface, then  $\Sigma$  is the set of all subareas that can be defined on the surface, and  $n(\mathcal{A}) = \int_{x \in \mathcal{A}} dn(x)$  is the detected photon number over the area  $\mathcal{A}$ . For any vectoral function  $h : \mathcal{X} \rightarrow \mathbb{R}^q$  on the detector space,

$$\check{h}(n) = \int h(x)dn(x), \quad (2.1)$$

a linear functional of  $n$ , is a random variable with statistics

$$\mathbb{E}(\check{h}) = \int h(x)d\bar{n}(x) = \nu(h), \quad (2.2)$$

$$\mathbb{V}(\check{h}) = \mathbb{E}(\check{h}\check{h}^\top) - \mathbb{E}(\check{h})\mathbb{E}(\check{h}^\top) = \nu(hh^\top), \quad (2.3)$$

where  $\mathbb{E}$  denotes the statistical expectation,  $\mathbb{V}$  denotes the covariance,  $\nu$  denotes the average with respect to the intensity measure  $\bar{n}$ ,  $\top$  denotes the matrix transpose, and all vectors in this paper are column vectors.

Suppose that  $\bar{n}$  depends on an unknown vectoral parameter  $\theta \in \Theta \subset \mathbb{R}^p$  with  $p$  entries and has a density  $f(x|\theta)$  with respect to a dominating measure  $\mu$  such that  $f(x|\theta) = d\bar{n}(x|\theta)/d\mu(x)$ . The log-likelihood derivatives are given by [62]

$$\check{S}_j(n|\theta) = \int \frac{\partial}{\partial \theta_j} \ln f(x|\theta)dn(x) - \frac{\partial}{\partial \theta_j} \int d\bar{n}(x|\theta). \quad (2.4)$$

As  $\check{S}$  is a linear functional of  $n$ , its covariance, called the Fisher information matrix, can be simplified via Eq. (2.3) and is given by [62]

$$J = \mathbb{V}(\check{S}) = \int S(x)[S(x)]^\top d\bar{n}(x) = \nu(SS^\top), \quad (2.5)$$

where  $S$  is a vector of detector-space functions given by

$$S_j(x|\theta) = \frac{\partial}{\partial \theta_j} \ln f(x|\theta). \quad (2.6)$$

Here  $\mathbb{V}$ ,  $\check{S}$ ,  $\bar{n}$ ,  $S$ , and  $\nu$  are all evaluated at the same  $\theta$ , and I assume hereafter that all functions of  $\theta$  are evaluated implicitly at the same  $\theta$ . Each  $S_j$  is hereafter called a score function, borrowing the same terminology for  $\check{S}$  in statistics [21,22]. An important distinction is that, whereas  $\mathbb{E}(\check{S}_j) = 0$ ,  $\nu(S_j)$  does not have to be zero, since  $\bar{n}$  does not have to be normalized.

Let  $\beta(\theta)$  be a scalar parameter of interest. If  $\beta(\theta) = \theta_k$ , for example, then all the other parameters in  $\theta$  are called nuisance parameters. For any unbiased estimator  $\check{\beta}(n)$ , the CRB on its variance is [5]

$$\mathbb{V}(\check{\beta}) \geq u^\top J^{-1}u = \text{CRB}, \quad u_j = \frac{\partial \beta}{\partial \theta_j}. \quad (2.7)$$

$J^{-1}$  seems intractable if  $\theta$  is infinite dimensional. The next section introduces a clever method.

## III. SEMIPARAMETRIC CRAMÉR-RAO BOUND

The key to the semiparametric theory is to treat random variables as elements in a Hilbert space [21,22]. Here I introduce another Hilbert space for detector-space functions on top of the statistical one for the purpose of computing the CRB for Poisson processes. Define an inner product between two scalar functions  $h_1, h_2 : \mathcal{X} \rightarrow \mathbb{R}$  as

$$(h_1, h_2) = \nu(h_1 h_2) = \int h_1(x)h_2(x)d\bar{n}(x), \quad (3.1)$$

and the norm as

$$\|h\| = \sqrt{(h, h)} = \sqrt{\nu(h^2)}. \quad (3.2)$$

With the inner product, a Hilbert space  $\mathcal{H}$  can be defined as the set of all square-summable functions, viz.,

$$\mathcal{H} = \{h(x) : \nu(h^2) < \infty\}. \quad (3.3)$$

Denote the set of score functions  $\{S_j\}$  as  $S$  in a slight abuse of notation. If the Fisher information  $J_{jj} = \nu(S_j^2) < \infty$  for all  $j$ ,  $S \in \mathcal{H}$ . Define the tangent space  $\mathcal{T} \subseteq \mathcal{H}$  of a parametric model as the linear span of  $S$ , or

$$\mathcal{T} = \{w^\top S : w \in \mathbb{R}^p\} = \text{span}(S). \quad (3.4)$$

Define also an ‘‘influence’’ function as any  $\check{\beta} \in \mathcal{H}$  that satisfies

$$\nu(\check{\beta}S) = u, \quad (3.5)$$

borrowing the name of a similar concept in statistics [21,22]. The Cauchy-Schwartz inequality  $\nu(\check{\beta}^2)[w^\top \nu(SS^\top)w] \geq (u^\top w)^2$  with  $w = [\nu(SS^\top)]^{-1}u$  then yields

$$\nu(\check{\beta}^2) \geq u^\top J^{-1}u, \quad (3.6)$$

the right-hand side of which coincides with the CRB given by Eq. (2.7). A way of evaluating the CRB is to consider the efficient influence, which is defined as the influence function that saturates Eq. (3.6), viz.,

$$\check{\beta}_{\text{eff}} = u^\top J^{-1}S = \nu(\check{\beta}S^\top)[\nu(SS^\top)]^{-1}S, \quad (3.7)$$

$$\text{CRB} = \nu(\check{\beta}_{\text{eff}}^2). \quad (3.8)$$

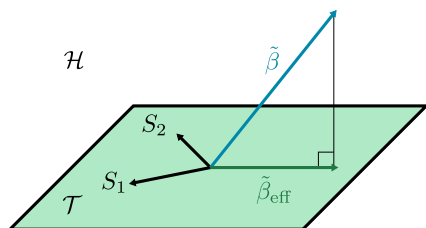


FIG. 1. The efficient influence  $\tilde{\beta}_{\text{eff}}$  is the orthogonal projection of any influence function  $\tilde{\beta} \in \mathcal{H}$  that satisfies Eq. (3.5) into the tangent space  $\mathcal{T} = \text{span}(S)$ . The norm of  $\tilde{\beta}_{\text{eff}}$  gives the CRB.

Equation (3.7) can be interpreted as the orthogonal projection of any influence function  $\tilde{\beta} \in \mathcal{H}$  that satisfies Eq. (3.5) into  $\mathcal{T}$ , viz.,

$$\tilde{\beta}_{\text{eff}} = \Pi(\tilde{\beta}|\mathcal{T}) = \arg \min_{h \in \mathcal{T}} \|\tilde{\beta} - h\|. \quad (3.9)$$

Figure 1 illustrates this concept.

Consider now the semiparametric scenario. For the purpose of this paper, it suffices to assume that the dimension of  $\theta$  is infinite but countable ( $p = \infty$ ). The score functions are still defined in the same way, but now there are infinitely many of them. The tangent space should be modified to be the closed linear span

$$\mathcal{T} = \overline{\text{span}}(S), \quad (3.10)$$

so that projection into it is well defined [63], and the semiparametric CRB is still given by Eqs. (3.5), (3.8), and (3.9); see Appendix A for a proof. This Hilbert-space approach is tractable when finding a candidate influence function is straightforward and the tangent space is so large that the candidate is already in it or at least very close to it. If the dimension of  $\theta$  is uncountable, the tangent space and the CRB can still be defined via the concept of parametric submodels [21,22], although it is not needed here.

If  $\beta$  can be expressed as a functional  $\beta(f)$ , a useful way of finding an influence function is to consider a functional derivative of  $\beta(f)$  with respect to  $h(x)$  defined as

$$\dot{\beta}(f, h) = \lim_{\epsilon \rightarrow 0} \frac{\beta((1 + \epsilon h)f) - \beta(f)}{\epsilon} \quad (3.11)$$

$$= \int \tilde{\beta}(x)h(x)f(x)d\mu(x) = v(\tilde{\beta}h), \quad (3.12)$$

which leads to

$$\frac{\partial \beta}{\partial \theta_j} = \lim_{\epsilon \rightarrow 0} \frac{\beta(f + \epsilon \partial f / \partial \theta_j) - \beta(f)}{\epsilon} \quad (3.13)$$

$$= \dot{\beta}(f, S_j) = v(\tilde{\beta}S_j) = u_j, \quad (3.14)$$

and the  $\tilde{\beta}(x)$  function obtained from the functional derivative is an influence function that satisfies Eq. (3.5). The simplest example is the linear functional

$$\beta(f) = \int \tilde{\beta}(x)f(x)d\mu(x) = v(\tilde{\beta}), \quad (3.15)$$

and  $\tilde{\beta}(x)$  is an influence function.

If the tangent space is so large that  $\mathcal{T} = \mathcal{H}$ , then a square-summable influence function is already in  $\mathcal{H} = \mathcal{T}$  and therefore efficient. There are often some constraints that make  $\mathcal{T}$

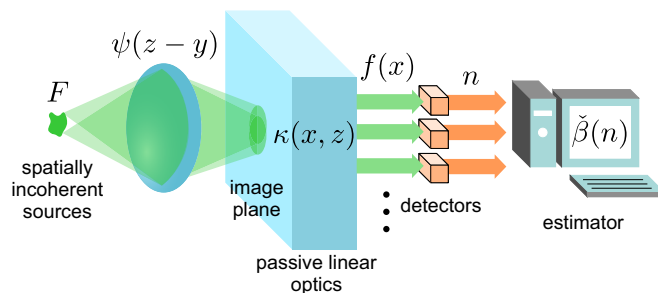


FIG. 2. A far-field incoherent imaging system. See the main text for definitions.

smaller, however, and the CRB is reduced as a result. For example, if the constraint can be expressed as

$$\gamma(f) = 0, \quad (3.16)$$

and its functional derivative is

$$\dot{\gamma}(f, h) = v(h\tilde{\gamma}), \quad (3.17)$$

then

$$\frac{\partial \gamma(f)}{\partial \theta_j} = \dot{\gamma}(f, S_j) = v(\tilde{\gamma}S_j) = \langle \tilde{\gamma}, S_j \rangle = 0, \quad (3.18)$$

and it follows that  $\tilde{\gamma}$  should be placed in the set that spans  $\mathcal{T}^\perp$ , the orthocomplement of  $\mathcal{T}$  in  $\mathcal{H}$ . In terms of  $\mathcal{T}^\perp$ , the efficient influence can be evaluated as

$$\tilde{\beta}_{\text{eff}} = \tilde{\beta} - \Pi(\tilde{\beta}|\mathcal{T}^\perp). \quad (3.19)$$

If  $\mathcal{T}^\perp = \text{span}(R)$ , then

$$\Pi(\tilde{\beta}|\mathcal{T}^\perp) = v(\tilde{\beta}R^\top)[v(RR^\top)]^{-1}R, \quad (3.20)$$

which is still tractable if  $R$  has a low dimension.

#### IV. INCOHERENT OPTICAL IMAGING

Consider a distribution of spatially incoherent sources described by the measure  $F$  on the object plane with coordinate  $y$ , a far-field paraxial imaging system with point-spread function  $\psi(z - y)$  for the field [2], further processing of the field on the image plane with coordinate  $z$  via passive linear optics with Green's function  $\kappa(x, z)$ , and Poisson noise at the output detectors labeled by  $x \in \mathcal{X}$ , as depicted by Fig. 2. For simplicity, assume one-dimensional imaging such that  $y, z \in \mathbb{R}$ ; generalization for two-dimensional imaging is possible [24,25] but not very interesting. The intensity can be described by the mixture model [4,23–25,30]

$$f(x) = \int \left| \int \kappa(x, z)\psi(z - y)dz \right|^2 dF(y), \quad (4.1)$$

where the image-plane coordinate  $z$  is normalized with respect to the magnification factor, both  $y$  and  $z$  are normalized with respect to the width of the point-spread function such that they are dimensionless, and  $\psi$  is normalized as  $\int |\psi(x)|^2 dx = 1$ . This semiclassical Poisson model can be derived from standard quantum optics [23,52,64].

For direct imaging with infinitesimal pixels,  $\kappa(x, z) = \sqrt{\tau}\delta(x - z)$ , where  $\tau$  is a positive conversion factor,

$x \in \mathcal{X} = \mathbb{R}$  denotes the position of each pixel,  $d\mu(x) = dx$ , and the image intensity obeys the convolution model

$$f(x) = \int H(x - y)dF(y), \quad H(x) = \tau|\psi(x)|^2, \quad (4.2)$$

which will be studied in Sec. V.

The most remarkable physics of the problem lies in the possibility of improving the measurement via optics with a different Green's function  $\kappa$ . Quantum information theory has shown that substantial improvement is possible for sub-diffraction objects, and SPADE has been found to be quantum optimal in many special cases [23–53]. In one version of SPADE,  $\kappa^*(q, z)$  is the  $q$ th mode function in the point-spread-function-adapted (PAD) basis [25,36], such that the output intensity is given by

$$f(q) = \int H(q|y)dF(y), \quad q \in \mathcal{X} = \mathbb{N}_0, \quad (4.3)$$

$$H(q|y) = \left| \int \kappa(q, z)\psi(z - y)dz \right|^2, \quad (4.4)$$

where  $\mu$  is simply the counting measure and  $\kappa$  and  $H$  obey special properties, as further discussed in Sec. VII. For a fair comparison, the quantum efficiencies of direct imaging and SPADE are assumed to be the same, meaning that [25]

$$\sum_{q=0}^{\infty} H(q|y) = \tau, \quad (4.5)$$

where  $\tau$  is the same factor as that for direct imaging. Then the expected photon number received in total, given by

$$N = \mathbb{E}[n(\mathcal{X})] = \bar{n}(\mathcal{X}) = \nu(1) = \tau \int dF(y), \quad (4.6)$$

is also the same.

### V. MOMENT ESTIMATION WITH DIRECT IMAGING

Consider the direct-imaging model given by Eq. (4.2). Assume that  $H$  can be expanded in a Taylor series as

$$H(x - y) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \frac{\partial^j H(x)}{\partial x^j} y^j, \quad (5.1)$$

which leads to

$$f(x) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \frac{\partial^j H(x)}{\partial x^j} \theta_j, \quad (5.2)$$

where the unknown parameters are the object moments defined by

$$\theta_j = \int y^j dF(y), \quad j \in \mathbb{N}_0. \quad (5.3)$$

For the CRB to hold, the parameter space should correspond to the condition that  $F$  contains an infinite number of point sources with different positions, as discussed in Appendix B. The score function for each  $\theta_j$  is

$$S_j(x) = \frac{(-1)^j}{j! f(x)} \frac{\partial^j H(x)}{\partial x^j}. \quad (5.4)$$

It turns out that the tangent space  $\mathcal{T}$  for this problem is equal to the whole Hilbert space  $\mathcal{H}$  under certain technical conditions, as shown in Appendix C.

Let the parameter of interest be

$$\beta = u^\top \theta = \sum_{j=0}^{\infty} u_j \theta_j, \quad (5.5)$$

where  $u$  is independent of  $\theta$ . To find a candidate influence function, a trick [65] is to consider the image moments  $\phi = (\phi_0, \phi_1, \dots)^\top$  given by

$$\phi_j = \int \tilde{\phi}_j(x) d\bar{n}(x) = \nu(\tilde{\phi}_j), \quad (5.6)$$

where

$$\tilde{\phi}_j(x) = x^j, \quad j \in \mathbb{N}_0 \quad (5.7)$$

are the monomials. Assuming that all the moments of  $F$  and  $H$  are finite such that all the moments of  $f$  are also finite,  $\phi$  can be related to the object moments via

$$\phi_j = \iint x^j H(x - y) dF(y) dx \quad (5.8)$$

$$= \iint (z + y)^j H(z) dF(y) dz \quad (5.9)$$

$$= \iint \sum_{k=0}^j \binom{j}{k} z^{j-k} y^k H(z) dF(y) dz \quad (5.10)$$

$$= \sum_{k=0}^{\infty} C_{jk} \theta_k, \quad (5.11)$$

where

$$C_{jk} = 1_{j \geq k} \binom{j}{k} \int H(x) x^{j-k} dx \quad (5.12)$$

and

$$1_{\text{proposition}} = \begin{cases} 1 & \text{if proposition is true,} \\ 0 & \text{otherwise.} \end{cases} \quad (5.13)$$

$C$  is a lower-triangular matrix, and with  $C_{jj} = \int H(x) dx = \tau > 0$ ,  $C^{-1}$  is well defined and also lower triangular even if the dimension of  $C$  is infinite, as shown in Appendix D. The object moments can then be related to the image moments by

$$\theta = C^{-1} \phi, \quad (5.14)$$

and  $\beta$  can be expressed as

$$\beta = u^\top \theta = u^\top C^{-1} \phi = \nu(u^\top C^{-1} \tilde{\phi}). \quad (5.15)$$

According to Eq. (3.15), an influence function is

$$\tilde{\beta}(x) = u^\top C^{-1} \tilde{\phi}(x) = u^\top \tilde{\theta}(x), \quad (5.16)$$

$$\tilde{\theta}(x) = C^{-1} \tilde{\phi}(x). \quad (5.17)$$

Since  $\mathcal{T} = \mathcal{H}$  as shown in Appendix C, the  $\tilde{\beta}$  given by Eq. (5.16) belongs to  $\mathcal{H} = \mathcal{T}$  and is efficient according to Eq. (3.9) as long as it is square summable. For example, if  $u$  contains a finite number of nonzero entries,  $\tilde{\beta}$  is a polynomial of  $x$  and must be square summable, since all the moments of

$f$  are assumed to be finite. The CRB is hence

$$\begin{aligned} \text{CRB}^{\text{(direct)}} &= v(\tilde{\beta}^2) = u^\top v(\tilde{\theta}\tilde{\theta}^\top)u \\ &= u^\top C^{-1}v(\tilde{\phi}\tilde{\phi}^\top)C^{-\top}u, \end{aligned} \quad (5.18)$$

where  $C^{-\top} = (C^{-1})^\top$ . This result coincides with that derived in Ref. [25] via a more direct but less rigorous method, which is repeated in Appendix E for completeness.

An unbiased and efficient estimator  $\check{\beta}(n)$  in terms of the observed process  $n$  can be constructed from the efficient influence as

$$\check{\beta}(n) = \int \tilde{\beta}(x)dn(x) = u^\top \check{\theta}(n), \quad (5.19)$$

where the object moment estimator is

$$\check{\theta}(n) = C^{-1}\check{\phi}(n), \quad \check{\phi}(n) = \int \tilde{\phi}(x)dn(x). \quad (5.20)$$

$\check{\beta}(n)$  is a linear filter of  $n$ , so its variance is  $\mathbb{V}(\check{\beta}) = v(\check{\beta}^2)$ , which coincides with the CRB. It is important to note that this estimator does not require any knowledge of the unknown parameters, as  $\check{\phi}(n)$  is simply the empirical moments of the observed image, and  $C^{-1}$  is a lower-triangular matrix that depends on the moments of the point-spread function  $H$ . The estimator still works even if the object happens to consist of a finite number of point sources and  $\theta$  is on the boundary of the parameter space, although its efficiency in that case is a more difficult question, as explained in Appendix B. Unlike some of the previous studies on superresolution [17–20], the results here place no restriction on the separations of the point sources and also account for Poisson noise explicitly.

## VI. CONSTRAINED CRAMÉR-RAO BOUND

In imaging, the parameters of interest are often the moments with respect to a normalized object distribution with  $\int dF(y) = 1$ . A simple way of modeling this is to assume that  $\theta_0 = 1$  is known. This constraint also makes the model comparable to those in Refs. [24,26,52]. Then,

$$N = \phi_0 = \tau\theta_0 \quad (6.1)$$

is known as well, implying the constraint  $\gamma(f) = \int f(x)dx - N = 0$ . The constraint can be differentiated to yield  $\dot{\gamma}(f, S_j) = v(S_j) = \langle S_j, 1 \rangle = 0$ , leading to  $\mathcal{T}^\perp = \text{span}(1)$ . The new efficient influence, according to Eqs. (3.19) and (3.20), should therefore be

$$\tilde{\beta}_{\text{eff}} = \tilde{\beta} - \Pi(\tilde{\beta}|\mathcal{T}^\perp) = \tilde{\beta} - \frac{v(\tilde{\beta})}{v(1)} = \tilde{\beta} - \frac{\beta}{N}. \quad (6.2)$$

The constrained CRB is now

$$\text{CRB}_{\theta_0=1}^{\text{(direct)}} = v(\tilde{\beta}_{\text{eff}}^2) = \frac{1}{N} [v_0(\tilde{\beta}_0^2) - \beta^2], \quad (6.3)$$

$$\tilde{\beta}_0(x) = N\tilde{\beta}(x) = u^\top (C/\tau)^{-1}\tilde{\phi}(x), \quad (6.4)$$

where  $v_0(h) = v(h)/v(1)$  is the normalized version of  $v$ . The CRB is necessarily lowered by the constraint. Other approaches to the constrained CRB yield the same result, as discussed in Appendix F.

To construct a near-efficient estimator, suppose that  $n(\mathcal{X}) = \int dn(x) = L > 0$  photons have been detected.

Then  $dn(x) = \sum_{i=1}^L 1_{x=X_i}$ , and the photon positions  $\{X_1, X_2, \dots, X_L\}$  are independent and identically distributed according to the probability measure  $\bar{n}/N$ . Consider the estimator

$$\check{\beta}(n) = \frac{1}{L} \int \tilde{\beta}_0(x)dn(x) = \frac{1}{L} \sum_{i=1}^L \tilde{\beta}_0(X_i). \quad (6.5)$$

It is straightforward to show that

$$\mathbb{E}(\check{\beta}|n(\mathcal{X}) = L) = v_0(\tilde{\beta}_0) = \beta, \quad (6.6)$$

$$\mathbb{V}(\check{\beta}|n(\mathcal{X}) = L) = \frac{1}{L} [v_0(\tilde{\beta}_0^2) - \beta^2], \quad (6.7)$$

which is close to the CRB given by Eq. (6.3) if  $L$  is close to its expected value  $N$ . The results are then consistent with standard results in semiparametric estimation concerning the moments of a normalized probability measure [21].

## VII. EVEN-MOMENT ESTIMATION WITH SPADE

Now consider the SPADE model given by Eqs. (4.3) and (4.4) and the Fourier transforms

$$\Psi(k) = \frac{1}{\sqrt{2\pi}} \int \psi(z) \exp(-ikz) dz, \quad (7.1)$$

$$\Phi_q(k) = \frac{1}{\sqrt{2\pi}} \int \kappa^*(q, z) \exp(-ikz) dz. \quad (7.2)$$

Suppose that  $\Phi = \{\Phi_q(k)\}$  is the PAD basis [25,36] given by

$$\Phi_q(k) = \sqrt{\tau} b_q(k) \Psi(k), \quad q \in \mathbb{N}_0, \quad (7.3)$$

where  $b = \{b_q(k) : q \in \mathbb{N}_0\}$  is the set of orthonormal polynomials defined by

$$\int |\Psi(k)|^2 b_q(k) b_r(k) dk = \delta_{qr}. \quad (7.4)$$

The polynomials exist for all  $q \in \mathbb{N}_0$  as long as the support of  $|\Psi(k)|^2$  is infinite [66], and the orthonormality of  $\Phi$  ensures that the measurement can be implemented by passive linear optics [23,30,36]. Equation (4.4) becomes

$$H(q|y) = \tau \left| \int |\Psi(k)|^2 b_q(k) \exp(-iky) dk \right|^2 \quad (7.5)$$

$$= \tau \left| \int |\Psi(k)|^2 b_q(k) \sum_{j=0}^{\infty} \frac{(-iky)^j}{j!} dk \right|^2. \quad (7.6)$$

As the  $b$  polynomials are derived by applying the Gram-Schmidt procedure to the monomials  $(1, k, k^2, \dots)^\top$ , their basic properties include  $\int |\Psi(k)|^2 b_q(k) k^r dk = 0$  if  $r < q$ ,  $\int |\Psi(k)|^2 b_q(k) k^q dk \neq 0$ , and  $b_q(k) = (-1)^q b_q(-k)$  if  $|\Psi(k)|^2$  is even, as is often the case in optics. These properties lead to

$$H(q|y) = \sum_{j=0}^{\infty} C_{qj} y^{2j}, \quad (7.7)$$

where  $C$  is an upper-triangular matrix ( $C_{qj} = 0$  if  $j < q$ ) with positive diagonal entries ( $C_{qq} > 0$ ). Equation (4.3) becomes

$$f(q) = \sum_{j=0}^{\infty} C_{qj} \theta_{2j}, \quad (7.8)$$

which depends on the even moments

$$\theta_{2j} = \int y^{2j} dF(y), \quad j \in \mathbb{N}_0. \quad (7.9)$$

The score function with respect to each  $\theta_{2j}$  becomes

$$S_j(q) = \frac{1}{f(q)} \frac{\partial f(q)}{\partial \theta_{2j}} = \frac{C_{qj}}{f(q)}. \quad (7.10)$$

Appendix G proves that  $\mathcal{T} = \overline{\text{span}}(S) = \mathcal{H}$ .

To find a candidate influence function, suppose that Eq. (7.8) can be inverted to give

$$\theta_{2j} = \sum_{q=0}^{\infty} (C^{-1})_{jq} f(q). \quad (7.11)$$

An influence function for  $\beta = u^\top \theta$  according to Eq. (3.15) is therefore

$$\tilde{\beta}(q) = u^\top \tilde{\theta}(q), \quad \tilde{\theta}_{2j}(q) = (C^{-1})_{jq}. \quad (7.12)$$

Since  $\mathcal{T} = \mathcal{H}$ , this  $\tilde{\beta}$  belongs to  $\mathcal{T}$  and is efficient as long as it is square summable. The CRB is hence

$$\begin{aligned} \text{CRB}^{(\text{SPADE})} &= v(\tilde{\beta}^2) = u^\top v(\tilde{\theta} \tilde{\theta}^\top) u \\ &= u^\top C^{-1} D C^{-\top} u, \end{aligned} \quad (7.13)$$

$$D_{jk} = f(j) \delta_{jk}. \quad (7.14)$$

A more direct but heuristic way of deriving Eq. (7.13) is shown in Appendix H. An unbiased and efficient estimator in terms of the observed detector counts  $n$  is

$$\check{\beta}(n) = \sum_{q=0}^{\infty} \tilde{\beta}(q) n(q) = u^\top \sum_{q=0}^{\infty} \tilde{\theta}(q) n(q). \quad (7.15)$$

This estimator has a variance  $\mathbb{V}(\check{\beta}) = v(\check{\beta}^2) = \text{CRB}^{(\text{SPADE})}$ , requires no knowledge of any unknown parameter, and still works even if the object happens to consist of a finite number of point sources, with no restriction on their separations. If  $\theta_0 = 1$ , the constrained CRB can be derived in ways similar to Sec. VI and Appendix F.

To estimate the odd moments of  $F$  via SPADE, variations of the PAD basis are needed [24,25]. The model is much more complicated and a derivation of the CRB and the efficient estimator is too tedious to work out here, but the upshot is the same: it can be shown that the tangent space for the problem encompasses the whole Hilbert space  $\mathcal{H}$ , the efficient influence can be retrieved from the relation  $\beta = v(\check{\beta})$ , and an unbiased and efficient estimator is  $\hat{\beta}(n) = \int \tilde{\beta}(x) dn(x)$ .

### A. Gaussian point-spread function

More explicit results can be obtained and the assumptions can be checked more carefully by assuming the Gaussian

point-spread function

$$\psi(z) = \frac{1}{(2\pi)^{1/4}} \exp\left(-\frac{z^2}{4}\right). \quad (7.16)$$

The PAD basis becomes the Hermite-Gaussian basis, and it can be shown that [23,24,57]

$$H(q|y) = \tau \exp\left(-\frac{y^2}{4}\right) \frac{(y/2)^{2q}}{q!}. \quad (7.17)$$

The  $C$  matrix in Eq. (7.7) can be determined by expanding  $\exp(-y^2/4) = \sum_{j=0}^{\infty} (-y^2/4)^j / j!$ , giving

$$C_{qj} = 1_{j \geq q} \frac{\tau (-1)^{j-q}}{4^j q! (j-q)!}. \quad (7.18)$$

It is not difficult to check that the matrix inverse of  $C$  is

$$\tilde{\theta}_{2j}(q) = (C^{-1})_{jq} = 1_{q \geq j} \frac{4^j q!}{\tau (q-j)!}, \quad (7.19)$$

which is a degree- $j$  polynomial of  $q$ .  $\sum_{q=0}^{\infty} \tilde{\theta}_{2j}(q) f(q)$  is the  $j$ th factorial moment of  $f$  and indeed equal to  $\theta_{2j}$ , since  $H(q|y)$  is Poisson and its factorial moment is  $\sum_{q=0}^{\infty} \tilde{\theta}_{2j}(q) H(q|y) = y^{2j}$  [67]. In general, each degree- $j$  moment of  $H(q|y)$  is a degree- $j$  polynomial of  $y^2$ , so each degree- $j$  moment of  $f(q)$  is a linear combination of the moments of  $F$  up to degree  $2j$ . All the moments of  $f$  are therefore finite as long as all the moments of  $F$  are finite. If  $u$  has a finite number of nonzero entries, the influence function given by Eqs. (7.12) is a polynomial of  $q$ , so  $v(\check{\beta}^2) < \infty$ , and  $\check{\beta} \in \mathcal{H}$  is ensured.

### B. Bandlimited point-spread function

Another important example is the bandlimited point-spread function given by

$$\Psi(k) = \frac{1_{|k| < 1}}{\sqrt{2}}. \quad (7.20)$$

$b$  is then the well-known set of Legendre polynomials [68]. Appendix I shows the detailed calculations; here I list the results only. Equation (7.5) becomes

$$H(q|y) = \tau (2q+1) j_q^2(y), \quad (7.21)$$

where  $j_q(y)$  is the spherical Bessel function of the first kind [68, Eq. (10.47.3)]. An influence function for estimating  $\theta_{2j}$  with  $\theta_{2j} = v(\check{\theta}_{2j})$  is

$$\tilde{\theta}_{2j}(q) = 1_{q \geq j} \frac{(2j+1)!! (2j-1)!!}{\tau} \binom{q+j}{2j}, \quad (7.22)$$

where  $!!$  denotes the double factorial [68].  $\tilde{\theta}_{2j}(q)$  is a degree- $2j$  polynomial of  $q$ , so  $\check{\beta}(q)$  is also a polynomial of  $q$  if  $u$  contains a finite number of nonzero entries. As long as all the moments of  $F$  are finite, all the moments of  $f$  can also be shown to be finite, and  $v(\check{\beta}^2) < \infty$  is ensured.

Notice that the direct-imaging theory in Sec. V breaks down for this bandlimited point-spread function, as the second and higher even moments of  $H(x) = \tau |\psi(x)|^2 = (\tau/\pi) \text{sinc}^2(x)$  are all infinite. The CRB in that case remains an open problem, although it is possible to apodize the point-spread function optically such that all its moments become

finite and the semiparametric estimator given by Eq. (5.19) has a finite variance. For example, if

$$\Psi(k) \propto 1_{|k|<1} \exp\left(-\frac{1}{k^2-1}\right), \quad (7.23)$$

then  $\Psi(k)$  is infinitely differentiable despite the hard bandwidth limit [69] and all the moments of  $|\psi(x)|^2$  are finite [25].

### VIII. COMPARISON OF IMAGING METHODS

The advantage of SPADE over direct imaging occurs in the subdiffraction regime, where the width  $\Delta$  of the object distribution  $F$  with respect to the origin is much smaller than the width of the point-spread function  $\psi$  [24–26,52]. As the width of  $\psi$  is normalized as 1, the regime is defined as

$$\Delta \ll 1, \quad (8.1)$$

and the object moments scale as

$$\theta_j = \theta_0 O(\Delta^j). \quad (8.2)$$

With the attainable CRBs given by Eqs. (5.18) and (7.13) at hand, I can now compare direct imaging and SPADE on the same semiparametric footing. Consider the estimation of a specific moment  $\theta_k$  with

$$u_j = \delta_{jk}. \quad (8.3)$$

For direct imaging in the subdiffraction regime, the image becomes close to the point-spread function, viz.,

$$f(x) \approx \theta_0 H(x) = N |\psi(x)|^2, \quad (8.4)$$

where  $N$ , the expected photon number received in total, is given by Eq. (4.6). With  $C_{jk} = \tau O(1)$  and  $v(\tilde{\phi}\tilde{\phi}^\top) = NO(1)$ , the CRB becomes

$$\text{CRB}^{(\text{direct})} = \frac{\theta_0^2}{N} O(1). \quad (8.5)$$

For SPADE on the other hand, notice that the  $C$  and  $C^{-1}$  matrices are upper-triangular, meaning that each  $f(q)$  depends only on  $\theta_{2q}$  and higher-order moments, implying

$$f(q) = NO(\Delta^{2q}). \quad (8.6)$$

The CRB for estimating  $\theta_k$ , where  $k$  is even, becomes

$$\text{CRB}^{(\text{SPADE})} = \frac{\theta_0^2}{N} O(\Delta^k), \quad (8.7)$$

which is much lower than Eq. (8.5) when  $\Delta \ll 1$  and  $k \geq 2$ . This is consistent with earlier approximate results [24,25]. An intuitive explanation of the enhancement, as well as a discussion of the limitations of SPADE, can be found in Ref. [52]. The constrained CRB with  $\theta_0 = 1$  becomes  $[O(\Delta^k) - \theta_k^2]/N = O(\Delta^k)/N$ , which is on the same order of magnitude as the fundamental quantum limit [26].

More exact and pleasing results can be obtained if  $\psi$  is Gaussian and given by Eq. (7.16). Consider for example the estimation of the second moment  $\theta_2$ . For direct imaging, it can be shown that

$$\text{CRB}^{(\text{direct})} = \frac{1}{\tau} (2\theta_0 + 4\theta_2 + \theta_4) = \frac{\theta_0^2}{N} O(1). \quad (8.8)$$

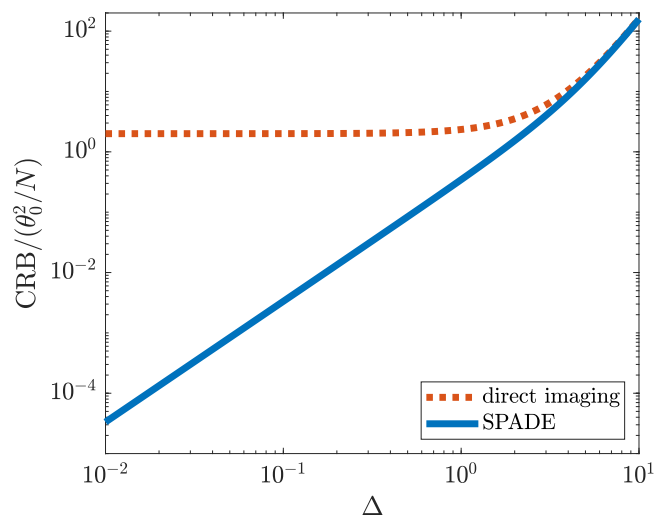


FIG. 3. The semiparametric CRBs for the second moment  $\theta_2$  given by Eqs. (8.8) and (8.9) vs the object size  $\Delta$  in log-log scale, if the point-spread function is the Gaussian given by Eq. (7.16) and the object happens to be the flat top given by Eq. (8.10). Both the CRBs and  $\Delta$  are normalized such that they are dimensionless.

For SPADE, on the other hand,

$$\text{CRB}^{(\text{SPADE})} = \frac{1}{\tau} (4\theta_2 + \theta_4) = \frac{\theta_0^2}{N} O(\Delta^2), \quad (8.9)$$

which not only beats direct imaging by a significant amount in the subdiffraction regime but is in fact superior for all parameter values. To further illustrate the difference between the two methods, suppose that the object happens to be a flat top given by

$$dF(y) = \frac{\theta_0}{\Delta} 1_{|y|<\Delta/2} dy. \quad (8.10)$$

Do note that the semiparametric CRBs do not assume the knowledge of this object shape, which is specified here only for the purpose of plotting examples of the CRBs. With  $\theta_2 = \theta_0 \Delta^2/12$  and  $\theta_4 = \theta_0 \Delta^4/80$ , Fig. 3 plots Eqs. (8.8) and (8.9) against  $\Delta$  in log-log scale. The gap between the two curves in the  $\Delta \ll 1$  regime is striking.

With the constraint  $\theta_0 = 1$ , the CRBs become

$$\text{CRB}_{\theta_0=1}^{(\text{direct})} = \frac{1}{N} (2 + 4\theta_2 + \theta_4 - \theta_2^2) = \frac{O(1)}{N}, \quad (8.11)$$

$$\text{CRB}_{\theta_0=1}^{(\text{SPADE})} = \frac{1}{N} (4\theta_2 + \theta_4 - \theta_2^2) = \frac{O(\Delta^2)}{N}. \quad (8.12)$$

It is noteworthy that Eq. (8.12) is exactly equal to the quantum limit given by Ref. [52, Eq. (E15)], meaning that SPADE is exactly quantum optimal—at all parameter values—for estimating the second moment. This is consistent with previous results concerning the estimation of two-point separation [23] and object size [24,28], but note that the previous results assume that the object shape is known, whereas the CRBs and the estimators here assume the opposite.

**IX. CONCLUSION**

The semiparametric theory set forth solves an important and difficult problem in incoherent optical imaging: the evaluation of the CRB and the efficient estimation of object parameters when little prior information about the object is available. The theory gives exact and achievable semiparametric CRBs for both direct imaging and SPADE, establishing the superiority and versatility of SPADE beyond the special parametric scenarios considered by previous studies.

Despite the elegant results, the theory has a few shortcomings. On the mathematical side, the conditions for the theory to hold seem difficult to check in the case of direct imaging with a non-Gaussian point-spread function, especially when the point-spread function has infinite moments. It is an open question whether this is merely a technicality or a hint at a whole new regime of statistics. On the practical side, the theory may be accused of assuming ideal conditions for both measurements, such as infinitesimal pixels for direct imaging, the availability of infinitely many modes for SPADE, perfect specification and knowledge of the optical systems, and the absence of excess noise. Reality is necessarily uglier, but the results here remain relevant by serving as fundamental limits (via the data-processing inequality [52,70]) and offering insights into the essential physics and statistics. The theoretical and experimental progress on SPADE and related methods so far [23–60,71–73] has provided encouragement that the theory has realistic potential, and the general results here should motivate further research into the wider applications of quantum-inspired imaging methods.

An interesting future direction is to generalize the semiparametric formalism for quantum estimation [74,75]. By treating the symmetric logarithmic derivatives of the quantum state  $\rho$  as the scores in the  $\mathcal{L}_h^2(\rho)$  space proposed by Holevo [75] and adopting a geometric picture [76], a quantum generalization of the semiparametric CRB can be envisioned. Whether it can solve any important problem, however, such as the quantum limit to incoherent imaging [26,27], remains to be seen.

**ACKNOWLEDGMENT**

This work is supported by the Singapore National Research Foundation under Project No. QEP-P7.

**APPENDIX A: PROOF OF THE SEMIPARAMETRIC CRB FOR POISSON PROCESSES**

Define the inner product between two random variables  $\check{r}_1$  and  $\check{r}_2$  as

$$\langle \check{r}_1, \check{r}_2 \rangle = \mathbb{E}(\check{r}_1 \check{r}_2), \tag{A1}$$

and the norm as

$$\|\check{r}\| = \sqrt{\langle \check{r}, \check{r} \rangle} = \sqrt{\mathbb{E}(\check{r}^2)}. \tag{A2}$$

Let the Hilbert space of zero-mean random variables be

$$\check{\mathcal{R}} = \{\check{r} : \mathbb{E}(\check{r}) = 0, \mathbb{E}(\check{r}^2) < \infty\}, \tag{A3}$$

and define

$$\check{\mathcal{T}} = \overline{\text{span}}(\check{\mathcal{S}}) \subseteq \check{\mathcal{R}}, \tag{A4}$$

where  $\check{\mathcal{S}}$  is defined by Eq. (2.4). Let  $\check{\delta} \in \check{\mathcal{R}}$  be any random variable that satisfies

$$\mathbb{E}(\check{\delta}\check{\mathcal{S}}) = u. \tag{A5}$$

The semiparametric CRB is [21,22]

$$\mathbb{E}(\check{\delta}^2) \geq \text{CRB} = \mathbb{E}(\check{\delta}_{\text{eff}}^2), \tag{A6}$$

$$\check{\delta}_{\text{eff}} = \Pi(\check{\delta}|\check{\mathcal{T}}) = \arg \min_{\check{h} \in \check{\mathcal{T}}} \|\check{\delta} - \check{h}\|. \tag{A7}$$

The proof can be done via a Pythagorean theorem [22] without recourse to the Cauchy-Schwartz inequality or the existence of  $J^{-1}$ .  $\check{\mathcal{S}}_j$  is called a score and  $\check{\delta}$  an influence in statistics [21,22]; this paper uses the same terminology for  $S$  and  $\check{\beta}$  in light of their resemblance to the statistical quantities.

The resemblance can be turned into a precise correspondence for a Poisson random measure by considering the subspace  $\check{\mathcal{H}} \subset \check{\mathcal{R}}$  of random variables that are linear with respect to  $n$ . Any element  $\check{h} \in \check{\mathcal{H}}$  can be expressed as

$$\check{h} = Uh = \int h(x)[dn(x) - d\bar{n}(x)], \tag{A8}$$

where  $U$  is a surjective linear map from  $\mathcal{H}$  to  $\check{\mathcal{H}}$ . Since

$$\langle Uh_1, Uh_2 \rangle = \langle h_1, h_2 \rangle \quad \forall h_1, h_2 \in \mathcal{H} \tag{A9}$$

by virtue of Eq. (2.3),  $\check{\mathcal{H}}$  is isomorphic to  $\mathcal{H}$  and  $U$  is unitary [63], and since  $\check{\mathcal{T}} \subseteq \check{\mathcal{H}}$  and  $\check{\mathcal{S}} = US$ ,  $\check{\mathcal{T}}$  is isomorphic to  $\mathcal{T}$ . Picking a  $\check{\delta} \in \check{\mathcal{H}}$  with

$$\check{\delta} = U\check{\beta} = \int \check{\beta}(x)[dn(x) - d\bar{n}(x)] \tag{A10}$$

leads to

$$\mathbb{E}(\check{\delta}\check{\mathcal{S}}) = \nu(\check{\beta}S) = u, \quad \check{\delta}_{\text{eff}} = U\check{\beta}_{\text{eff}}, \tag{A11}$$

where  $\check{\beta}_{\text{eff}}$  is given by Eq. (3.9) because of Eq. (A7) and the isomorphisms. The CRB becomes

$$\text{CRB} = \mathbb{E}(\check{\delta}_{\text{eff}}^2) = \nu(\check{\beta}_{\text{eff}}^2), \tag{A12}$$

which is Eq. (3.8).

**APPENDIX B: THE MOMENT PARAMETER SPACE**

Define an  $s \times s$  Hankel matrix with respect to a real-number sequence  $\theta = (\theta_0, \theta_1, \dots)^T$  as

$$M_{jk}^{(s)}(\theta) = \theta_{j+k}, \quad j, k \in \{0, 1, \dots, s-1\}. \tag{B1}$$

If  $\theta$  is a moment sequence that arises from a nonnegative measure  $F$ ,

$$w^T M^{(s)} w = \int \left( \sum_{j=0}^{s-1} w_j y^j \right)^2 dF(y) \tag{B2}$$

is nonnegative for any real vector  $w$ , and all Hankel matrices are positive semidefinite, viz.,

$$M^{(s)} \geq 0 \quad \forall s \in \mathbb{N}_1. \tag{B3}$$

Conversely, any sequence with Hankel matrices that obey Eq. (B3) can be expressed in the form of Eq. (5.3) with a nonnegative  $F$  by virtue of Hamburger’s theorem [77].



For the CRB to hold for a  $p$ -dimensional  $\theta$ , the parameter space  $\Theta$  should be an open subset of  $\mathbb{R}^p$  [70,78]. Intuitively, the requirement makes sense because all the parameters in  $\theta$  are unknown and  $\theta$  should be allowed to vary in a neighborhood; otherwise, the problem would be overparametrized. If  $\Theta$  is not an open subset, the parameter space would be constrained and the CRB may be modified [78]. When all the moments are unknown parameters, consider the set

$$\Theta = \{\theta : M^{(s)}(\theta) > 0 \forall s \in \mathbb{N}_1\}. \quad (\text{B4})$$

Each  $\theta \in \Theta$  corresponds to a measure with infinite support size  $r = \infty$  [77]. The proof can be done by observing that the polynomial in Eq. (B2) has at most  $s-1$  zeros and the integral is strictly positive for any  $w \neq 0$  if and only if  $r \geq s$ , and therefore the constraint for  $\Theta$  is satisfied if and only if  $r = \infty$ . For  $r < \infty$ ,  $F$  can be expressed in terms of its support  $\{Y_j : 0 \leq j \leq r-1, Y_j < Y_{j+1}\}$  and weights  $\{F_j > 0 : 0 \leq j \leq r-1\}$  as

$$dF(y) = \sum_{j=0}^{r-1} F_j \delta(y - Y_j), \quad \frac{dF(y)}{dy} = \sum_{j=0}^{r-1} F_j \delta'(y - Y_j). \quad (\text{B5})$$

In the context of optics,  $r$  is the minimum number of point sources that can describe the object distribution. The assumption of Eq. (B4) as the parameter space is consistent with the infinite-support assumption for semiparametric estimation with mixture models [21, Sec. 6.5], and it also makes intuitive sense, at least as a necessary condition—with  $r$  point sources, there are only  $2r$  unknown parameters, and the problem would be overparametrized if all the moments are assumed to be unknown. Further inequality constraints on  $\theta$  may be needed to ensure the convergence of the Taylor series in Eqs. (5.2) and (7.8), although they should not affect the CRB as long as  $\theta$  obeys and stays away from them [78].

The boundary of  $\Theta$  corresponds to measures with finite support size  $r < \infty$ . If  $s \leq r$ , then  $M^{(s)} > 0$  and  $M^{(s)}$  is full rank (rank =  $s$ ), but if  $s > r$ , I can write

$$M^{(s)} = V^\top \text{diag}(F)V, \quad (\text{B6})$$

$$V_{jk} = Y_j^k, \quad \text{diag}(F)_{jk} = 1_{0 \leq j \leq r-1} F_j \delta_{jk}, \quad (\text{B7})$$

where  $V$  and  $\text{diag}(F)$  are both  $s \times s$  matrices.  $V$  is the Vandermonde matrix and invertible since  $\{Y_j : 0 \leq j \leq s-1\}$ , which should match the support of  $F$  for the first  $r$  elements, can be assumed to be distinct [79]. With  $M^{(s)} \geq 0$  and  $\text{diag}(F) \geq 0$ , Sylvester's law of inertia [79] implies that the rank of  $M^{(s)}$  is the same as the rank of  $\text{diag}(F)$ , which is  $r$ . In other words, the rank of  $M^{(s)}$  is  $\min(r, s)$ , and any finite  $r$  means that  $M^{(s)}$  is rank deficient and does not satisfy the strict  $M^{(s)} > 0$  for all  $s > r$ . Whether the CRB still holds for  $\theta$  on the boundary is a difficult question, considering that the parameter space here is infinite dimensional and it is not obvious how existing finite-dimensional results regarding the CRB on a boundary [78] can be applied.

### APPENDIX C: TANGENT SPACE FOR THE DIRECT-IMAGING MODEL

Consider the tangent space  $\mathcal{T}$  given by Eq. (3.10) and the score functions given by Eq. (5.4) for direct imaging. First

note that  $S \subset \mathcal{H}$ , as the Fisher information  $J_{jj} = \langle S_j, S_j \rangle = \nu(S_j^2)$  is assumed to be finite for all  $j$ . Recent calculations in quantum estimation theory suggest that this assumption is reasonable for any measurement [26]. To prove  $\mathcal{T} = \overline{\text{span}}(S) = \mathcal{H}$ , the standard method is to show that the only element in  $\mathcal{H}$  orthogonal to  $S$  is 0 [63]; that is,

$$\langle h, S_j \rangle = 0 \quad \forall j \in \mathbb{N}_0 \quad (\text{C1})$$

implies  $h = 0$  (almost everywhere with respect to  $\bar{n}$ ). Here, I list a few approaches with various levels of rigor.

The first approach is to consider the set of orthogonal polynomials

$$a = \left\{ a_j(x) = \sum_{k=0}^{\infty} A_{jk} \tilde{\phi}_k(x) : j \in \mathbb{N}_0, \langle a_j, a_k \rangle = \delta_{jk} \right\}, \quad (\text{C2})$$

where  $A$  is a lower triangular matrix with nonzero diagonal entries and can be determined by applying the Gram-Schmidt procedure to the monomials  $\tilde{\phi}_k(x)$  given by Eq. (5.7) [66]. Under rather general conditions on  $f$ , the polynomials form an orthonormal basis of  $\mathcal{H}$  [66], viz.,

$$\mathcal{H} = \overline{\text{span}}(a), \quad (\text{C3})$$

and I can write Eq. (C1) as

$$\langle h, S_j \rangle = \sum_{k=0}^{\infty} \langle h, a_k \rangle \langle a_k, S_j \rangle = 0 \quad \forall j \in \mathbb{N}_0, \quad (\text{C4})$$

or, more compactly,

$$B^\top w = 0, \quad w_k = \langle h, a_k \rangle, \quad B_{kj} = \langle a_k, S_j \rangle. \quad (\text{C5})$$

If the only solution to Eq. (C5) is  $w = 0$ , then the only solution to Eq. (C4) is  $h = 0$ . This is equivalent to the condition that  $B^\top$  is injective.

Integration by parts yields

$$B_{kj} = \frac{(-1)^j}{j!} \int a_k(x) \frac{\partial^j H(x)}{\partial x^j} dx = \sum_{l=0}^{\infty} A_{kl} C_{lj}, \quad (\text{C6})$$

where  $C$  is the same as Eq. (5.12). Since both  $A$  and  $C$  are lower triangular with nonzero diagonal entries,  $B = AC$  is also lower triangular with nonzero diagonal entries, and  $B^\top$  has a well-defined matrix inverse  $(B^\top)^{-1} = (B^{-1})^\top = A^{-\top} C^{-\top}$  in the sense that

$$B^\top (B^\top)^{-1} = (B^\top)^{-1} B^\top = I, \quad (\text{C7})$$

where  $I$  is the identity matrix; see Appendix D for details. If the matrices were finite-dimensional, the existence of a matrix inverse would imply

$$(B^\top)^{-1} (B^\top w) = [(B^\top)^{-1} B^\top] w = w, \quad (\text{C8})$$

and the only solution to Eq. (C5) would be  $w = 0$ . This proof is not entirely satisfactory, however, as Eq. (C8) assumes that the product of the infinite-dimensional matrices is associative. Associativity assumes that the order of the sums involved in the matrix product can be interchanged, but it cannot be guaranteed for infinite-dimensional matrices. In other words, the existence of a matrix inverse for  $B^\top$  may not imply that  $B^\top$  is injective.

A more rigorous approach is to define

$$\chi_y(x) = \sum_{j=0}^{\infty} y^j S_j(x), \quad y \in \mathcal{Y} \subset \mathbb{R}, \quad (\text{C9})$$

and notice that Eq. (C1) implies

$$\langle h, \chi_y \rangle = \int h(x) \sum_{j=0}^{\infty} y^j \frac{(-1)^j}{j!} \frac{\partial^j H(x)}{\partial x^j} dx \quad (\text{C10})$$

$$= \int h(x) H(x - y) dx = 0 \quad \forall y \in \mathcal{Y}. \quad (\text{C11})$$

The unique solution to Eq. (C11) is  $h = 0$  if the family  $\{H(x - y) : y \in \mathcal{Y}\}$  satisfies a property called completeness in statistics [5]. For example, if  $H$  is Gaussian,  $\{H\}$  is a full-rank exponential family for any open subset  $\mathcal{Y} \subset \mathbb{R}$  and therefore complete [5]. An even more rigorous formulation of this approach [21] is to treat  $\langle h, \chi_y \rangle$  as an operator that maps  $h \in \mathcal{H}$  to a function of  $y$  in another Hilbert space and then show that the null space of the operator consists of only  $h = 0$ . The proof again boils down to the requirement that  $\{H\}$  should be complete; see Ref. [21, Sec. 6.5].

#### APPENDIX D: INVERSE OF AN INFINITE-DIMENSIONAL TRIANGULAR MATRIX

Let  $C$  be an infinite-dimensional matrix with entries indexed by  $(j, k) \in \mathbb{N}_0^2$ . Define its formal matrix inverse  $C^{-1}$  as another infinite-dimensional matrix that satisfies

$$\sum_{l=0}^{\infty} C_{jl} (C^{-1})_{lk} = \delta_{jk}. \quad (\text{D1})$$

If  $C$  is lower triangular with nonzero diagonal entries, viz.,

$$C_{jk} = 0 \text{ if } k > j, \quad C_{jj} \neq 0, \quad (\text{D2})$$

then  $C^{-1}$  can be found by a recursive relation as follows. Define  $C^{(s)}$  as the  $s \times s$  upper-left submatrix of  $C$ . Write  $C^{(s+1)}$  and  $(C^{-1})^{(s+1)}$  as the partitions

$$C^{(s+1)} = \begin{pmatrix} C^{(s)} & 0 \\ c^\top & C_{ss} \end{pmatrix}, \quad (\text{D3})$$

$$(C^{-1})^{(s+1)} = \begin{pmatrix} (C^{-1})^{(s)} & 0 \\ d^\top & (C^{-1})_{ss} \end{pmatrix}. \quad (\text{D4})$$

Given  $(C^{-1})^{(s)} = (C^{(s)})^{-1}$ ,

$$d^\top = -\frac{c^\top (C^{(s)})^{-1}}{C_{ss}}, \quad (C^{-1})_{ss} = \frac{1}{C_{ss}}, \quad (\text{D5})$$

and the recursion starts from  $(C^{-1})^{(1)} = (C^{(1)})^{-1}$  with one element  $(C^{-1})_{00} = 1/C_{00}$ . The matrix inverse of an infinite-dimensional upper-triangular matrix can be defined in a similar way.

Although the product of infinite-dimensional matrices may not be associative, it can still be proved by induction that  $D(Cu) = (DC)u$  for any vector  $u$  if  $D$  and  $C$  are lower triangular, because

$$D(Cu) = \sum_{k=0}^{\infty} D_{jk} \sum_{l=0}^{\infty} C_{kl} u_l = \sum_{k=0}^j D_{jk} \sum_{l=0}^k C_{kl} u_l \quad (\text{D6})$$

involves finite sums only. Thus, it is safe to assume that  $C^{-1}(Cu) = (C^{-1}C)u = u$  and  $C(C^{-1}u) = (CC^{-1})u = u$  if  $C$  is lower triangular with nonzero diagonal entries.

#### APPENDIX E: AN ALTERNATIVE DERIVATION OF THE CRAMÉR-RAO BOUND FOR DIRECT IMAGING

Consider the problem described in Sec. V. Since the polynomials given by Eq. (C2) are an orthonormal basis, the information matrix for the moment parameters can be expressed as

$$J_{jk} = \langle S_j, S_k \rangle = \sum_{l=0}^{\infty} \langle S_j, a_l \rangle \langle a_l, S_k \rangle, \quad (\text{E1})$$

meaning that  $J = B^\top B$ , where  $B = AC$  is given by Eq. (C6). Ignoring the complications of multiplying infinite-dimensional matrices, the CRB becomes

$$J^{-1} = B^{-1} B^{-\top} = C^{-1} A^{-1} A^{-\top} C^{-\top}. \quad (\text{E2})$$

To evaluate  $A^{-1} A^{-\top}$ , notice that the orthonormality of  $a$  can be written as

$$\langle a_j, a_k \rangle = \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} A_{jl} \langle \tilde{\phi}_l, \tilde{\phi}_m \rangle A_{km} = \delta_{jk}, \quad (\text{E3})$$

where  $\tilde{\phi}$  is the monomials given by Eq. (5.7). In other words,

$$Av(\tilde{\phi}\tilde{\phi}^\top)A^\top = I, \quad A^{-1}A^{-\top} = v(\tilde{\phi}\tilde{\phi}^\top), \quad (\text{E4})$$

giving

$$J^{-1} = C^{-1} v(\tilde{\phi}\tilde{\phi}^\top) C^{-\top}. \quad (\text{E5})$$

This leads to Eq. (5.18) for the parameter  $\beta = u^\top \theta$ .

#### APPENDIX F: ALTERNATIVE APPROACHES TO THE CONSTRAINED CRAMÉR-RAO BOUND

One way of deriving the constrained CRB if  $\theta_0$  is known is to consider the information matrix  $\tilde{J}$  with respect to the parameters  $\vartheta = (\theta_1, \theta_2, \dots)^\top$  without  $\theta_0$ . Then  $\theta = (\theta_0, \vartheta^\top)^\top$ , and  $\tilde{J}$  can be written as the submatrix of  $J$ , or

$$J = \begin{pmatrix} J_{00} & j^\top \\ j & \tilde{J} \end{pmatrix}, \quad (\text{F1})$$

where  $j$  is a column vector. Ignore the complications of dealing with infinite-dimensional matrices and partition  $J^{-1}$  similarly as

$$J^{-1} = \begin{pmatrix} E_{00} & e^\top \\ e & \tilde{E} \end{pmatrix}. \quad (\text{F2})$$

Then, it is straightforward to show that

$$\tilde{J}^{-1} = \tilde{E} - \frac{ee^\top}{E_{00}}. \quad (\text{F3})$$

Let  $\tilde{\vartheta} = (\tilde{\theta}_1, \tilde{\theta}_2, \dots)^\top$ , and observe that  $\tilde{\theta}_0 = 1/C_{00}$  from Eqs. (5.17), (5.12), and (5.6). Then, Eq. (5.18) implies that

$$\tilde{E} = v(\tilde{\vartheta}\tilde{\vartheta}^\top), \quad (\text{F4})$$

$$e = v(\tilde{\vartheta}\tilde{\theta}_0) = \frac{v(\tilde{\vartheta})}{C_{00}} = \frac{\vartheta}{C_{00}}, \quad (\text{F5})$$

$$E_{00} = v(\tilde{\theta}_0\tilde{\theta}_0) = \frac{v(1)}{C_{00}^2} = \frac{\phi_0}{C_{00}^2}. \quad (\text{F6})$$

Hence,

$$\tilde{J}^{-1} = \nu(\tilde{\vartheta}\tilde{\vartheta}^\top) - \frac{\vartheta\vartheta^\top}{\phi_0}, \quad (\text{F7})$$

which implies Eq. (6.3) if the parameter of interest is defined as  $\beta = u^\top\theta$  with  $u_0 = 0$ .

Yet another way of deriving the constrained CRB can be found in Ref. [78], which can be shown to lead to the same result here.

#### APPENDIX G: TANGENT SPACE FOR THE SPADE MODEL

The proof is similar to the first approach in Appendix C. Consider  $\mathcal{H} = \overline{\text{span}}(a)$  in terms of an obvious orthonormal basis

$$a = \{a_j(q) = \delta_{jq}/\sqrt{f(j)} : j \in \mathbb{N}_0\}. \quad (\text{G1})$$

Any  $h \in \mathcal{H}$  orthogonal to the  $S$  given by Eq. (7.10) obeys

$$\langle h, S_j \rangle = \sum_{k=0}^{\infty} \langle h, a_k \rangle \langle a_k, S_j \rangle = 0 \quad \forall j \in \mathbb{N}_0, \quad (\text{G2})$$

which can be written as

$$B^\top w = 0, \quad w_k = \langle h, a_k \rangle, \quad (\text{G3})$$

$$B_{jk} = \langle a_j, S_k \rangle = \frac{C_{jk}}{\sqrt{f(j)}}. \quad (\text{G4})$$

As  $C$  is upper triangular with nonzero diagonal entries and  $f > 0$  is assumed,  $B^\top$  is lower triangular with nonzero diagonal entries, and induction can be used to prove that the only solution to  $B^\top w = 0$  is  $w = 0$ , or equivalently  $h = 0$ . Hence,  $\mathcal{T} = \overline{\text{span}}(S) = \mathcal{H}$ . The proof is easier than the one in Appendix C because  $B^\top$  here is lower triangular rather than upper triangular.

An alternative proof, similar to the second approach in Appendix C and Ref. [21, Sec. 6.5] but less fruitful in this case, is to define

$$\chi_y(x) = \sum_{j=0}^{\infty} y^{2j} S_j(x), \quad y \in \mathcal{Y} \subset \mathbb{R}, \quad (\text{G5})$$

consider

$$\langle h, \chi_y \rangle = \sum_{q=0}^{\infty} h(q) H(q|y) = 0, \quad (\text{G6})$$

and use the completeness of  $\{H(q|y) : y \in \mathcal{Y}\}$  to prove the unique solution  $h = 0$ . If  $H$  is Poisson, for example, then  $\{H\}$  is a full-rank exponential family and therefore complete [5].

#### APPENDIX H: AN ALTERNATIVE DERIVATION OF THE CRAMÉR-RAO BOUND FOR SPADE

Consider the problem described in Sec. VII. With the orthonormal basis given by Eq. (G1) and the  $B$  matrix given by Eq. (G4), the information matrix with respect to the moment parameters can again be expressed as  $J = B^\top B$  according to Eq. (E1). With Eq. (G4),  $B^{-1}$  becomes

$$(B^{-1})_{jq} = (C^{-1})_{jq} \sqrt{f(q)}. \quad (\text{H1})$$

Ignoring the complications of multiplying infinite-dimensional matrices, the CRB  $J^{-1} = B^{-1}B^{-\top}$  is

$$(J^{-1})_{jk} = \sum_{q=0}^{\infty} (C^{-1})_{jq} f(q) (C^{-1})_{kq} = C^{-1} D C^{-\top}, \quad (\text{H2})$$

where  $D$  is given by Eq. (7.14) and the CRB for  $\beta = u^\top\theta$  coincides with Eq. (7.13).

#### APPENDIX I: CALCULATIONS CONCERNING SPADE FOR A BANDLIMITED POINT-SPREAD FUNCTION

Consider the point-spread function given by Eq. (7.20). The standard Legendre polynomials are defined in terms of

$$\frac{1}{2} \int_{-1}^1 P_q(k) P_p(k) dk = \frac{1}{2q+1} \delta_{qp}, \quad (\text{I1})$$

such that the orthonormal version is

$$b_q(k) = \sqrt{2q+1} P_q(k). \quad (\text{I2})$$

The Fourier transform of the polynomial is [68, Eq. (18.17.19)]

$$\frac{1}{2} \int_{-1}^1 b_q(k) \exp(iky) dk = \sqrt{2q+1} j_q(y), \quad (\text{I3})$$

where  $j_q(y)$  is the spherical Bessel function of the first kind [68, Eq. (10.47.3)]. Then, Eq. (7.21) follows from Eqs. (7.5) and (I3).

Let

$$\tilde{H}(q|y) = \frac{H(q|y)}{\tau} = (2q+1) j_q^2(y). \quad (\text{I4})$$

From Ref. [68, Eq. (10.60.2)], one can derive the useful formula

$$\sum_{q=0}^{\infty} \tilde{H}(q|y) P_q(k) = \text{sinc } w = \begin{cases} (\sin w)/w, & w \neq 0, \\ 1, & w = 0, \end{cases} \quad (\text{I5})$$

$$w = y\sqrt{2-2k}. \quad (\text{I6})$$

For example, since  $P_q(1) = 1$ , one can check that  $\sum_{q=0}^{\infty} \tilde{H}(q|y) = 1$  in accordance with Eq. (4.5). Using the facts

$$\text{sinc } w = \frac{1}{2} \int_{-1}^1 \exp(iwz) dz = \sum_{l=0}^{\infty} \frac{(-1)^l w^{2l}}{(2l+1)!}, \quad (\text{I7})$$

$$\frac{dw}{dk} = -\frac{y}{w}, \quad P_q^{(j)}(1) = \left. \frac{d^j P_q(k)}{dk^j} \right|_{k=1}, \quad (\text{I8})$$

it can also be shown that

$$\sum_{q=0}^{\infty} \tilde{H}(q|y) P_q^{(j)}(1) = \left. \frac{d^j \text{sinc } w}{dk^j} \right|_{k=1} = \frac{y^{2j}}{(2j+1)!!}, \quad (\text{I9})$$

which leads to

$$\sum_{q=0}^{\infty} f(q) P_q^{(j)}(1) = \frac{\tau \theta_{2j}}{(2j+1)!!}. \quad (\text{I10})$$

An influence function for estimating  $\theta_{2j}$  is hence

$$\tilde{\theta}_{2j}(q) = \frac{(2j+1)!!}{\tau} P_q^{(j)}(1), \quad (I11)$$

which obeys  $\theta_{2j} = v(\tilde{\theta}_{2j})$ . To derive an explicit expression for  $P_q^{(j)}(1)$ , consider the Rodrigues formula [68, Eq. (14.7.13)]

$$P_q(k) = \frac{1}{2^q q!} \frac{d^q}{dk^q} (k^2 - 1)^q, \quad (I12)$$

which leads to

$$P_q(k) = \sum_{l=0}^q \binom{q}{l} \binom{q+l}{l} \left(\frac{k-1}{2}\right)^l, \quad (I13)$$

$$P_q^{(j)}(1) = 1_{q \geq j} (2j-1)!! \binom{q+j}{2j}, \quad (I14)$$

and Eq. (7.22) results.

To bound the moments of  $\tilde{H}$  and  $f$ , consider a lower bound on the binomial coefficient for  $j \geq 1$  given by [80, pp. 1186]

$$\binom{q+j}{2j} \geq \frac{(q+j)^{2j}}{(2j)^{2j}} \geq \frac{q^{2j}}{(2j)^{2j}}, \quad (I15)$$

such that each even moment of  $\tilde{H}$  can be bounded as

$$\sum_{q=0}^{\infty} \tilde{H}(q|y) q^{2j} = \sum_{q=0}^{j-1} \tilde{H}(q|y) q^{2j} + \sum_{q=j}^{\infty} \tilde{H}(q|y) q^{2j} \quad (I16)$$

$$\leq (j-1)^{2j} + \frac{(2j)^{2j}}{(2j-1)!!} \sum_{q=0}^{\infty} \tilde{H}(q|y) P_q^{(j)}(1) \quad (I17)$$

$$= (j-1)^{2j} + \frac{(2j)^{2j} y^{2j}}{(2j-1)!! (2j+1)!!}. \quad (I18)$$

This means that each even moment of  $f(q)$  is bounded as

$$v(q^{2j}) \leq \tau \left[ (j-1)^{2j} \theta_0 + \frac{(2j)^{2j} \theta_{2j}}{(2j-1)!! (2j+1)!!} \right]. \quad (I19)$$

With  $v(q^0) = v(1) = \tau \theta_0$ ,  $v(q^{2j}) < \infty$  for all  $j \in \mathbb{N}_0$  as long as  $\theta_0$  and  $\theta_{2j}$  are finite. Odd moments can be bounded via the Cauchy-Schwartz inequality  $[v(q^j)]^2 \leq v(1)v(q^{2j})$ . Hence, all the moments of  $f$  are finite as long as all the moments of  $F$  are finite.

---

[1] M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Cambridge University Press, Cambridge, UK, 1999).

[2] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 2004).

[3] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge University Press, Cambridge, UK, 1995).

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

[5] E. L. Lehmann and George Casella, *Theory of Point Estimation* (Springer, New York, 1998).

[6] E. J. Farrell, Information content of photoelectric star images, *J. Opt. Soc. Am.* **56**, 578 (1966).

[7] M.-J. Tsai and K.-P. Dunn, Performance limitations on parameter estimation of closely spaced optical targets using shot-noise detector model, Tech. Rep. ADA073462, Lincoln Laboratory, MIT, Cambridge, MA, 1979.

[8] J. Zmuidzinas, Cramér-Rao sensitivity limits for astronomical instruments: Implications for interferometer design, *J. Opt. Soc. Am. A* **20**, 218 (2003).

[9] E. D. Feigelson and G. J. Babu, *Modern Statistical Methods for Astronomy* (Cambridge University Press, Cambridge, UK, 2012).

[10] S. Ram, E. S. Ward, and R. J. Ober, Beyond Rayleigh's criterion: A resolution measure with application to single-molecule microscopy, *Proc. Nat. Acad. Sci. USA* **103**, 4457 (2006).

[11] A. Small and S. Stahlheber, Fluorophore localization algorithms for super-resolution microscopy, *Nat. Methods* **11**, 267 (2014).

[12] H. Deschout, F. C. Zanacchi, M. Mlodzianoski, A. Diaspro, J. Bewersdorff, S. T. Hess, and K. Braeckmans, Precisely and accurately localizing single emitters in fluorescence microscopy, *Nat. Methods* **11**, 253 (2014).

[13] J. Chao, E. S. Ward, and R. J. Ober, Fisher information theory for parameter estimation in single molecule microscopy: Tutorial, *J. Opt. Soc. Am. A* **33**, B36 (2016).

[14] A. von Diezmann, Y. Shechtman, and W. E. Moerner, Three-dimensional localization of single molecules for super-resolution imaging and single-particle tracking, *Chem. Rev.* **117**, 7244 (2017).

[15] E. Bettens, D. Van Dyck, A. J. den Dekker, J. Sijbers, and A. van den Bos, Model-based two-object resolution from observations having counting statistics, *Ultramicroscopy* **77**, 37 (1999).

[16] S. Van Aert, A. J. den Dekker, D. Van Dyck, and A. van den Bos, High-resolution electron microscopy and electron tomography: Resolution versus precision, *J. Struct. Biol.* **138**, 21 (2002).

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

[18] E. J. Candès and C. Fernandez-Granda, Super-resolution from noisy data, *J. Fourier Anal. Appl.* **19**, 1229 (2013).

[19] E. J. Candès and C. Fernandez-Granda, Towards a mathematical theory of super-resolution, *Commun. Pure Appl. Math.* **67**, 906 (2014).

[20] G. Schiebinger, E. Robeva, and B. Recht, Superresolution without separation, *Inf. Inference: J. IMA* **7**, 1 (2018).

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

[22] A. A. Tsiatis, *Semiparametric Theory and Missing Data* (Springer, New York, 2006).

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

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

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

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

- [27] S. Zhou and L. Jiang, Modern description of Rayleigh's criterion, *Phys. Rev. A* **99**, 013808 (2019).
- [28] 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).
- [29] R. Nair and M. Tsang, Interferometric superlocalization of two incoherent optical point sources, *Opt. Express* **24**, 3684 (2016).
- [30] M. Tsang, R. Nair, and X.-M. Lu, Quantum information for semiclassical optics, *Proc. SPIE* **10029**, 1002903 (2016).
- [31] R. Nair and M. Tsang, Far-Field Superresolution of Thermal Electromagnetic Sources at the Quantum Limit, *Phys. Rev. Lett.* **117**, 190801 (2016).
- [32] C. Lupo and S. Pirandola, Ultimate Precision Bound of Quantum and Subwavelength Imaging, *Phys. Rev. Lett.* **117**, 190802 (2016).
- [33] M. Tsang, Conservative classical and quantum resolution limits for incoherent imaging, *J. Mod. Opt.* **65**, 1385 (2018).
- [34] S. Z. Ang, R. Nair, and M. Tsang, Quantum limit for two-dimensional resolution of two incoherent optical point sources, *Phys. Rev. A* **95**, 063847 (2017).
- [35] X.-M. Lu, H. Krovi, R. Nair, S. Guha, and J. H. Shapiro, Quantum-optimal detection of one-versus-two incoherent optical sources with arbitrary separation, *npj Quant. Inf.* **4**, 64 (2018).
- [36] 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).
- [37] F. Yang, R. Nair, M. Tsang, C. Simon, and A. I. Lvovsky, Fisher information for far-field linear optical superresolution via homodyne or heterodyne detection in a higher-order local oscillator mode, *Phys. Rev. A* **96**, 063829 (2017).
- [38] R. Kerviche, S. Guha, and A. Ashok, Fundamental limit of resolving two point sources limited by an arbitrary point spread function, in *2017 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Aachen, Germany, 2017), pp. 441–445.
- [39] A. Chrostowski, R. Demkowicz-Dobrzański, M. Jarzyna, and K. Banaszek, On super-resolution imaging as a multiparameter estimation problem, *Int. J. Quantum. Inform.* **15**, 1740005 (2017).
- [40] J. Řeháček, Z. Hradil, B. Stoklasa, M. Paúr, J. Grover, A. Krzic, and L. L. Sánchez-Soto, Multiparameter quantum metrology of incoherent point sources: Towards realistic superresolution, *Phys. Rev. A* **96**, 062107 (2017).
- [41] J. Řeháček, Z. Hradil, D. Koutný, J. Grover, A. Krzic, and L. L. Sánchez-Soto, Optimal measurements for quantum spatial superresolution, *Phys. Rev. A* **98**, 012103 (2018).
- [42] M. P. Backlund, Y. Shechtman, and R. L. Walsworth, Fundamental Precision Bounds For Three-Dimensional Optical Localization Microscopy with Poisson Statistics, *Phys. Rev. Lett.* **121**, 023904 (2018).
- [43] C. Napoli, S. Piano, R. Leach, G. Adesso, and T. Tufarelli, Towards Superresolution Surface Metrology: Quantum Estimation of Angular and Axial Separations, *Phys. Rev. Lett.* **122**, 140505 (2019).
- [44] Z. Yu and S. Prasad, Quantum Limited Superresolution of an Incoherent Source Pair in Three Dimensions, *Phys. Rev. Lett.* **121**, 180504 (2018).
- [45] S. Prasad and Z. Yu, Quantum-limited superlocalization and superresolution of a source pair in three dimensions, *Phys. Rev. A* **99**, 022116 (2019).
- [46] S. Prasad, Quantum limited super-resolution of an unequal-brightness source pair in three dimensions, [arXiv:1908.10949](https://arxiv.org/abs/1908.10949) [quant-ph].
- [47] W. Larson and B. E. A. Saleh, Resurgence of Rayleigh's curse in the presence of partial coherence, *Optica* **5**, 1382 (2018).
- [48] M. Tsang and R. Nair, Resurgence of Rayleigh's curse in the presence of partial coherence: Comment, *Optica* **6**, 400 (2019).
- [49] W. Larson and B. E. A. Saleh, Resurgence of Rayleigh's curse in the presence of partial coherence: Reply, *Optica* **6**, 402 (2019).
- [50] M. R. Grace, Z. Dutton, A. Ashok, and S. Guha, Approaching quantum limited super-resolution imaging without prior knowledge of the object location, [arXiv:1908.01996](https://arxiv.org/abs/1908.01996) [physics, physics:quant-ph].
- [51] 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).
- [52] M. Tsang, Resolving starlight: a quantum perspective, [arXiv:1906.02064](https://arxiv.org/abs/1906.02064) [astro-ph, physics:physics, physics:quant-ph].
- [53] E. Bisketzi, D. Branford, and A. Datta, Quantum limits of localisation microscopy, [arXiv:1907.11657](https://arxiv.org/abs/1907.11657) [physics, physics:quant-ph].
- [54] Z. S. Tang, K. Durak, and A. Ling, Fault-tolerant and finite-error localization for point emitters within the diffraction limit, *Opt. Express* **24**, 22004 (2016).
- [55] W.-K. Tham, H. Ferretti, and A. M. Steinberg, Beating Rayleigh's Curse by Imaging Using Phase Information, *Phys. Rev. Lett.* **118**, 070801 (2017).
- [56] M. Paúr, B. Stoklasa, Z. Hradil, L. L. Sánchez-Soto, and J. Řeháček, Achieving the ultimate optical resolution, *Optica* **3**, 1144 (2016).
- [57] 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).
- [58] J. M. Donohue, V. Ansari, J. Řeháček, Z. Hradil, B. Stoklasa, M. Paúr, L. L. Sánchez-Soto, and C. Silberhorn, Quantum-Limited Time-Frequency Estimation Through Mode-Selective Photon Measurement, *Phys. Rev. Lett.* **121**, 090501 (2018).
- [59] J. Hassett, T. Malhorta, M. A. Alonso, R. W. Boyd, S. M. H. Rafsanjani, and A. N. Vamivakas, Sub-Rayleigh limit localization with a spatial mode analyzer, in *Frontiers in Optics/Laser Science (2018)*, OSA Technical Digest (Optical Society of America, Washington, DC, 2018), p. JW4A.124.
- [60] Y. Zhou, J. Yang, J. D. Hassett, S. M. H. Rafsanjani, M. Mirhosseini, A. N. Vamivakas, A. N. Jordan, Z. Shi, and R. W. Boyd, Quantum-limited estimation of the axial separation of two incoherent point sources, *Optica* **6**, 534 (2019).
- [61] E. Çinlar, *Probability and Stochastics* (Springer, New York, 2011).
- [62] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space* (Springer-Verlag, New York, 1991).
- [63] M. Reed and B. Simon, *Methods of Modern Mathematical Physics. I: Functional Analysis* (Academic Press, San Diego, 1980).

- [64] M. Tsang, Quantum Nonlocality in Weak-Thermal-Light Interferometry, *Phys. Rev. Lett.* **107**, 270402 (2011).
- [65] A. Meister, Support estimation via moment estimation in presence of noise, *Statistics* **40**, 259 (2006).
- [66] C. F. Dunkl and Y. Xu, *Orthogonal Polynomials of Several Variables* (Cambridge University Press, Cambridge, UK, 2001).
- [67] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes* (Springer-Verlag, New York, 2003), Vol. 1.
- [68] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark (eds.), *NIST Handbook of Mathematical Functions* (NIST and Cambridge University Press, Cambridge, UK, 2010).
- [69] L. Debnath and F. A. Shah, *Wavelet Transforms and Their Applications* (Springer, New York, 2015).
- [70] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory* (Springer, New York, 1981).
- [71] M. Parniak, S. Borówka, K. Boroszko, W. Wasilewski, K. Banaszek, and R. Demkowicz-Dobrzański, Beating the Rayleigh Limit Using Two-Photon Interference, *Phys. Rev. Lett.* **121**, 250503 (2018).
- [72] 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).
- [73] 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).
- [74] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [75] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Edizioni della Normale, Pisa, Italy, 2011).
- [76] A. Fujiwara, Geometry of quantum estimation theory, in *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, edited by Masahito Hayashi (World Scientific, Singapore, 2005), Chap. 18, pp. 229–286.
- [77] K. Schmüdgen, *The Moment Problem* (Springer, Cham, 2017).
- [78] J. D. Gorman and A. O. Hero, Lower bounds for parametric estimation with constraints, *IEEE Trans. Inf. Theory* **36**, 1285 (1990).
- [79] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1985).
- [80] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* (MIT Press, Cambridge, MA, 2009).