THE EUROPEAN
PHYSICAL JOURNAL PLUS

Review

Optical resolution from Fisher information

L. Motka[1], B. Stoklasa[1], M. D'Angelo[2], P. Facchi[2,3], A. Garuccio[2,3], Z. Hradil[1], S. Pascazio[2,3], F.V. Pepe[3,4], Y.S. Teo[1], J. Řeháček[1], and L.L. Sánchez-Soto[5,6,a]

[1] Department of Optics, Palacký University, CZ-77146 Olomouc, Czech Republic
[2] Dipartimento di Fisica and MECENAS, Università di Bari, I-70126 Bari, Italy
[3] INFN, Sezione di Bari, I-70126 Bari, Italy
[4] Museo Storico della Fisica e Centro Studi e Ricerche "Enrico Fermi", I-00184 Roma, Italy
[5] Departamento de Óptica, Facultad de Física, Universidad Complutense, E-28040 Madrid, Spain
[6] Max-Planck-Institut für die Physik des Lichts, D-91058 Erlangen, Germany

Abstract. The information gained by performing a measurement on a physical system is most appropriately assessed by the Fisher information, which in fact establishes lower bounds on estimation errors for an arbitrary unbiased estimator. We revisit the basic properties of the Fisher information and demonstrate its potential to quantify the resolution of optical systems. We illustrate this with some conceptually important examples, such as single-slit diffraction, spectroscopy and superresolution techniques.

1 Introduction

One of the numerous definitions of the word "information" in the online Merriam-Webster dictionary [1] is "knowledge obtained from investigation, study, or instruction", typically about a certain subject or event. Nowadays, this notion has influenced the world of global communication in which we live. Other terms, such as "intelligence", "news", "facts", and "data" are often regarded as synonyms of this word. In the same dictionary entry, right after two other definitions that immediately follow, a slightly more technical meaning of this word reads "a numerical quantity that measures the uncertainty in the outcome of an experiment to be performed". This operational definition forms the basis of our understanding of Nature through data acquisition and statistical inference.

As with all scientific methods, it is necessary to determine the "accuracy" (or the "degree of conformity of a measure to a standard or a true value" [2]) of the information gained. Among related concepts, the accuracy of the acquired scientific information also correlates, in some respects, with its "exactness" and "correctness", thereby setting the standards of scientific truth.

In physics, any piece of information about a system is always masked by noise, which originates from several confounding factors, such as imperfections in the measurement apparatus, experimental limitations, and other plausible error sources. In optics, signal measurements of a given object yield statistical information about quantities attributed to it, which may be extracted using known methods of statistical inference and information theory. In general, the inferred values of physical quantities differ from the corresponding true values because of the presence of noise in the data. With more refined and optimized data analysis strategies, noise can be reduced to retrieve the original piece of information of interest. The uncertainties of these inferred values also depend on the choice of the statistical estimator for the measured parameters and on the particular configurations of the measurements. We shall employ a well-known statistical quantity—the Fisher information—to quantify the information gained *via a measurement with any optical device*. In this way, we will be able to evaluate and optimize the performance of measurement schemes and/or signal estimation strategies.

As exemplifying confirmation of its extensive compatibility with optics, the concept of Fisher information shall be applied to diffraction problems, spectroscopic measurements with noisy position-sensitive photodetectors and optical imaging of spatially bounded objects. These problems are intimately connected with important theoretical concepts, such as the point spread function, Rayleigh resolution limit and superresolution techniques.

---

[a] e-mail: lsanchez@fis.ucm.es
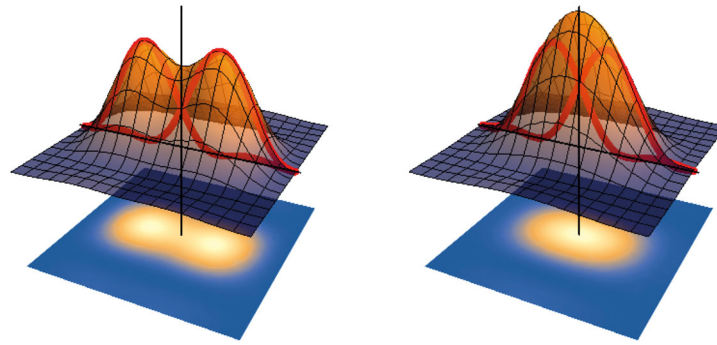
Fig. 1. Rayleigh criterion for incoherent imaging of two point sources. Surface plots of the image intensity distributions and density plots (planes below) are shown for two different separations of the two point objects. The red curves represent the single point intensities. The left (right) panel corresponds to a separation equal to (smaller than) the Rayleigh separation. Even though the intensity on the right shows only one maximum, the actual shape is broader than a single point intensity and contains information about two points. This information is retrievable by data analysis.

This paper is organized as follows. The standard theory of two-point resolution in optics is first briefly reviewed in sect. 2. This is followed by an introduction to the concept of the Fisher information in sect. 3. This machinery is applied in sect. 4 to some selected examples. Spectroscopic resolution is introduced in sect. 5 whereas sect. 6 discusses superresolution. Finally, sect. 7 serves as an epilogical treatise to biased estimation strategies that can potentially beat the standard inference limits defined by the Fisher information.

2 Two-point resolution of an imaging system

In optics, perfect imaging implies the existence of a one-to-one correspondence between points of the object plane and those of the image plane. Unfortunately, this perfect correspondence is prohibited by the laws of wave optics, where points are imaged in general as extended diffraction *spots*. The corresponding normalized intensity distribution is known as the point spread function (PSF). Put differently, the PSF represents the normalized response of an optical imaging system to a point light source. This function equivalently describes the image degradation due to diffraction phenomena and optical aberrations and provides a simple and intuitive way of evaluating the quality of an optical imaging system.

Assuming linearity and spatial invariance of an incoherent optical imaging system, the output intensity at the image plane, I_i, can be expressed as a convolution of the input intensity at the object plane, I_o, and the system PSF h:

$$I_i(x,y) = \int \mathrm{d}x' \int \mathrm{d}y' \, I_o(x',y')h(x-x', y-y'), \tag{1}$$

where (x,y) and (x',y') are the coordinates of the object and image planes, respectively. Neglecting optical aberrations and assuming the Fraunhofer approximation, the shape of the diffraction spot is readily expressed as the scaled Fourier transform of the aperture function of the imaging system. For example, for a rectangular aperture of dimensions $a \times b$, we get

$$h(x,y) = \mathrm{sinc}^2\left(\frac{\pi a x}{\lambda z}\right) \mathrm{sinc}^2\left(\frac{\pi b y}{\lambda z}\right), \tag{2}$$

where $\mathrm{sinc}\, x = (\sin x)/x$, λ is the central wavelength, and z is the image distance.

For a circular aperture of diameter a, the PSF shows cylindrical symmetry and can be expressed in terms of the Bessel function of first kind J_1, in polar coordinates $(r = \sqrt{x^2+y^2})$, as

$$h(r) = \left[2\frac{J_1(\pi a r/\lambda z)}{\pi a r/\lambda z}\right]^2. \tag{3}$$

Two-point resolution, which is defined as the ability to resolve two point sources of equal intensity, is a widely used measure of the overall resolving capabilities of an imaging system. The classical Rayleigh criterion [3] is certainly the most famous way to deal with the problem: according to this criterion, two point sources are just resolved if the maximum of the first PSF coincides with the first minimum of the second one. Two point sources are *optically resolved* if the separation is larger than that defined by the criterion. Figure 1 describes this pictorially. For instance, with an

optical system possessing a circular aperture of radius a, the Rayleigh criterion establishes a separation d between the sources of

$$d \gtrsim 1.22 \frac{\lambda z}{a} \,. \tag{4}$$

Nevertheless, this criterion should not be taken dogmatically [4]. Good visual resolution of two point sources, such as a double star, requires two separated maxima. Moreover, optimal digital processing of the data acquired by a camera allows, in principle, for an improvement of the optical resolution beyond the Rayleigh limit. Going back to fig. 1, upon comparing the imaging quality at the Rayleigh limit shown on the left, the image on the right, although poorly resolved according the Rayleigh criterion, evidently differs in shape and size from the image of a single point. Hence some information about any hidden features of the object is preserved (think of an object signal coming from two stars that are very close to each other) and can be retrieved with clever post processing. This measurement limit falls in the superresolution regime, that is beyond the standard resolution limit defined by the Rayleigh criterion. We shall discuss this in sect. 6, with more detailed statistical analysis.

3 Fisher information

Consider a measurement consisting of n output channels described by the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$, where $\boldsymbol{X} = \boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is one of the many possible data vectors obtainable from a given physical system. The probability $p(\boldsymbol{x}|\theta)$ for measuring $\boldsymbol{x}$ is a conditional probability that depends on the value of the unknown true parameter θ that describes some property of the system. A main goal of this article is to infer the value of the unknown parameter θ as faithfully as possible from the measured data $\boldsymbol{x}$—a well-known procedure of parameter estimation in statistics. We denote an estimator for θ calculated from such an estimation procedure by $\widehat{\theta}(\boldsymbol{x})$. Owing to the presence of noise, any such estimator $\widehat{\theta}(\boldsymbol{x})$ always possesses an inherent statistical uncertainty $\Delta\widehat{\theta}$, which depends on the choice of estimation strategy.

For example, upon tossing a coin N times, we observe a certain number x of "head"s. This measurement data can be used to estimate the true bias θ of the coin, that is, the probability of observing a "head" in a single flip. One plausible choice of estimator would be $\widehat{\theta}(x) = \widehat{\theta}_1(x) = x/N$—the relative frequency of observing head in the experiment. If one ignores the observed data x and insists that the coin is fair, then $\widehat{\theta}(x) = \widehat{\theta}_2(x) = 1/2$ is yet another option. Depending on the situation, either the former or the latter estimator shall more accurately describe the coin. When no additional prior information about the bias θ of the coin is known, $\widehat{\theta}_1(x)$ provides a much more reliable estimate as it converges to the true value in the limit of large N (a consistent estimator).

An important relation between the measurement statistics and estimation error is the *Cramér-Rao inequality* for the estimation *mean squared error* (MSE) of an unbiased estimator $\widehat{\theta}$, characterized by the property $\overline{\widehat{\theta}} = \theta$ [5,6], where the bar denotes average,

$$(\Delta\widehat{\theta})^2 \geq \frac{1}{F(\theta)} \,, \tag{5}$$

where

$$F(\theta) = \int \mathrm{d}\boldsymbol{x} \left(\frac{\partial}{\partial\theta} \log p(\boldsymbol{x}|\theta) \right)^2 p(\boldsymbol{x}|\theta) \tag{6}$$

is the Fisher information and $(\Delta\widehat{\theta})^2 = \overline{(\widehat{\theta} - \theta)^2}$ stands for the variance. Estimators saturating the *Cramér-Rao lower bound* (CRLB) $1/F(\theta)$ are called efficient estimators.

To proceed further we note that for n independent channels, the conditional probability takes the form
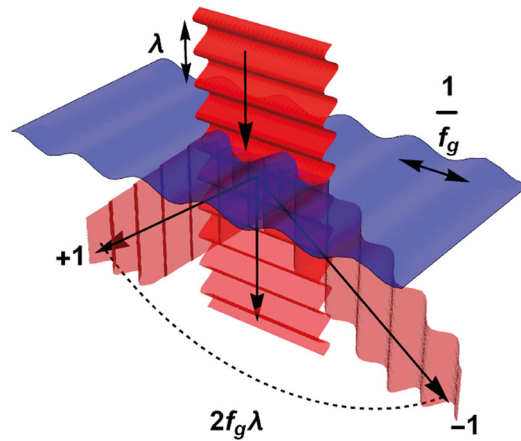
$$p(\boldsymbol{x}|\theta) = p(x_1, x_2, \ldots, x_n|\theta) = \prod_{j=1}^{n} p_j(x_j|\theta), \tag{7}$$

where $p_j(x_j|\theta)$ is the conditional probability for the j-th measurement output [7]. As an example, suppose that x_j is a continuous variable and

$$p_j(x_j|\theta) = \mathcal{N}\left(\bar{x}_j, (\Delta x_j)^2\right) = \frac{1}{\sqrt{2\pi(\Delta x_j)^2}} \exp\left[-\frac{(x_j - \bar{x}_j)^2}{2(\Delta x_j)^2} \right], \tag{8}$$

is a normal distribution with mean $\bar{x}_j = \bar{x}_j(\theta)$ and standard deviation Δx_j, independent of θ. Then, the Fisher information is

$$F(\theta) = \sum_{j=1}^{n} \frac{\bar{x}_j'^2}{(\Delta x_j)^2} \,, \tag{9}$$

Fig. 2. Diffraction of light through an amplitude grating. When light of wavelength λ illuminates the amplitude grating, part of its energy propagates in new subsequent directions that are represented by diffraction orders. Here the grating has a sinusoidal profile of amplitude transmission with period d and is plotted in blue. The picture shows the diffraction orders $m = +1$ and $m = -1$, where the angle between the two propagation directions is related to the grating period d and λ according to eq. (13).

where a prime denotes a derivative with respect to the unknown true parameter θ. The formula in eq. (9) is also valid for the Poissonian distribution

$$p_j(x_j = k|\theta) = \frac{\lambda_j^k \, e^{-\lambda_j}}{k!}, \tag{10}$$

with x_j taking random integer values k and $\lambda_j = \bar{x}_j$.

For another important class of measurements that involve sequential counting of sampling events, such that $\boldsymbol{x}$ is now a sequence of event occurrences collected into the vector $\boldsymbol{x} = (m_1, m_2, \ldots, m_n)$ of integers $\sum_{j=1}^{n} m_j = N$, and

$$p_j(x_j|\theta) = p_j(m_j|\theta) = p_j(\theta)^{m_j}, \tag{11}$$

the corresponding data statistics is a multinomial distribution and the following formula

$$F(\theta) = N \sum_{j=1}^{n} \frac{p_j'^{\,2}}{p_j} \tag{12}$$

holds. We shall not discuss here the relationship with the quantum aspects of quantum tomography and quantum Fisher information, that are analyzed in ref. [8].

4 Examples

4.1 Diffraction through an amplitude grating: estimating wavelength

For a better understanding of the role played by the Fisher information in optics, we begin by looking at a simple measurement of the wavelength of a laser source using an amplitude grating and a collimator. Behind the amplitude grating, diffraction spots appear, which are here represented as *diffraction orders* (see fig. 2). The mean angle of propagation ϕ_m of an optical beam after passing through the grating is a simple function of the diffraction order m:
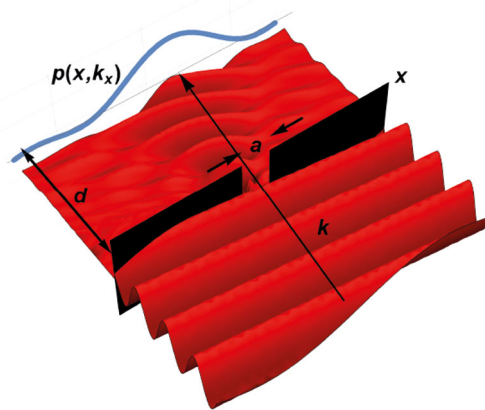
$$\phi_m = m f_{\mathrm{g}} \lambda, \tag{13}$$

with m integer and f_{g} being the grating frequency.

Let us consider the measurement of two output channels corresponding to diffraction orders $m = 1$ and $m = -1$. For large separation of the these spots, one may assume that these angular measurement outcomes are independent and normally distributed,

$$p(\phi_{\pm 1}|\lambda) = \mathcal{N}\left(\pm \lambda f_{\mathrm{g}}, (\Delta\phi)^2\right), \tag{14}$$

where $(\Delta\phi)^2$ is the corresponding angular uncertainty that characterizes the size of each diffraction spot (note that $\lambda f_{\mathrm{g}} \gg \Delta\phi$). In general, $(\Delta\phi)^2$ varies with λ, but for simplicity, we may take this to be a constant for small variations in λ, which is so for many practical purposes. The resulting joint probability for the angular outcomes is factorized:

$$p(\phi_1, \phi_{-1}|\lambda) = p(\phi_1|\lambda)\, p(\phi_{-1}|\lambda) = \mathcal{N}\left(\lambda f_{\mathrm{g}}, (\Delta\phi)^2\right) \mathcal{N}\left(-\lambda f_{\mathrm{g}}, (\Delta\phi)^2\right). \tag{15}$$

Fig. 3. Demonstration of the mutual uncertainty relation between the position and the wave vector. The electromagnetic field that is incident on the single slit subsequently propagates to the far-field zone behind the slit, where the position-sensitive photodetector is placed. The corresponding intensity pattern is measured here and consequently, the wave vector can be estimated from the collected data.

Every unbiased estimator $\widehat{\lambda}$ for the true wavelength λ will have an MSE $(\Delta\widehat{\lambda})^2$ and, according to eq. (9), this MSE is bounded from below by the inverse Fisher information:

$$(\Delta\widehat{\lambda})^2 \geq \frac{1}{F}, \qquad F = \frac{2f_{\mathrm{g}}^2}{(\Delta\phi)^2}. \tag{16}$$

4.2 Diffraction through a slit: estimating position

A different aspect of analyzing measurement uncertainties using the Fisher information can be demonstrated through the problem of single-slit diffraction. The measurement apparatus consists of the slit and a position-sensitive photodetector placed in the far-field zone behind the slit (see fig. 3). Since the position and wave vector $\boldsymbol{k} = (k_x, k_y, k_z)$ are conjugate variables, one expects their uncertainties to be mutually related. We suppose that z labels the axis of propagation and consider electromagnetic waves propagating from $z < 0$ towards a single slit of width a placed in the $z = 0$ plane, with field amplitude $E(x, y = 0, z) \propto \mathrm{e}^{\mathrm{i}(k_z z + k_x x)}$.

We make the additional assumption of paraxial propagation with respect to the z-axis; namely, $k_x \ll k_z$. If the slit is parallel to the y-axis and extends from $x = -a/2$ to $x = a/2$, the intensity distribution in x (for any y) in the $z = 0$ plane reads

$$p_0(x) = \frac{1}{a}\,\mathrm{rect}\left(\frac{x}{a}\right), \qquad \mathrm{rect}(x) = \begin{cases} 1, & \text{if } |x| < 1/2, \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

In the far-field limit, the field amplitude $\mathrm{e}^{\mathrm{i}k_x x}$ at $z = 0$ can be used to compute the intensity registered by the photodetector placed in the far-field plane at $z = d \gg a^2/\lambda$ using the Fraunhofer approximation [9]. The resulting intensity distribution is

$$p_d(x; k_x) = \frac{a k_z}{2\pi d}\,\mathrm{sinc}^2\,\xi, \qquad \xi = \frac{a}{2}\left(\frac{k_z x}{d} - k_x\right). \tag{18}$$

The joint parametric dependence on the position x and wave number k_x defines a relationship between the two physical quantities in wave propagation. The corresponding uncertainties Δx and Δk_x would also depend on this relationship. This means that when a measurement is made on these two quantities for the same light source, their uncertainties cannot vary arbitrarily.

The identity of this relationship between these uncertainties is revealed by simply calculating

$$(\Delta x)^2 = \int \mathrm{d}x\, x^2 p_0(x) = \frac{a^2}{12}, \qquad F(k_x) = \int \mathrm{d}x \left(\frac{\partial}{\partial k_x} \log p_d(x; k_x)\right)^2 p_d(x; k_x) = \frac{a^2}{3}, \tag{19}$$

so that, we find the uncertainty relation [10]

$$(\Delta x)^2 (\Delta k_x)^2 \geq \frac{(\Delta x)^2}{F(k_x)} = \frac{1}{4}. \tag{20}$$
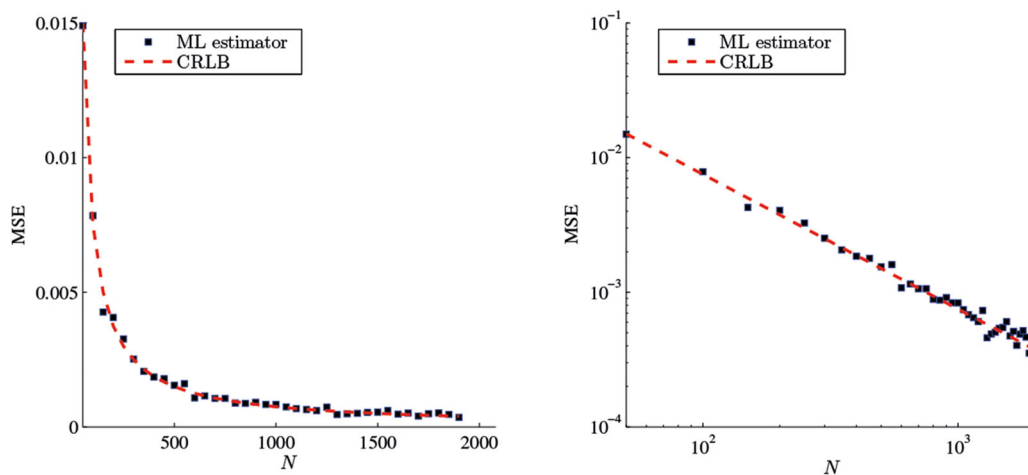
Fig. 4. The linear- and log-scale plots of the CRLB and the MSE of the ML estimator for the parameter k_x constructed out of a given set of position data of sample size N show that the MSE of the ML estimator effectively hits the CRLB of $3/(4N)$ [10]. For simplicity, all parameters are reduced to dimensionless quantities. The statistical fluctuation arises from calculating the MSE with data of 200 simulated runs for each value of N.

One might be tempted to believe that the inequality in (20) forms the basis of Heisenberg's uncertainty relation $(\Delta x)^2 (\Delta p_x)^2 \geq \hbar^2/4$ just by a simple substitution of $p_x = \hbar k_x$. However, the inequality (20) has a completely different origin than the Heisenberg relation [11–14]. The classical inequality quantifies the principal limitation in performing statistical inference on the wave number k_x. In particular, for a *fixed set of data*, the product on the left-hand side varies with different estimation strategies. For instance, for the *same* set of position data collected, one may either choose to take the sample mean to obtain a linear estimator $\widehat{k}_x$ for the unknown k_x that would give an infinite MSE as a property of the distribution function in eq. (18), or one may choose the Maximum Likelihood (ML) estimator, which will saturate this inequality (see fig. 4). Uncertainty relations based on the CRLB, such as that in (20), are extremely useful for evaluating the performance of complicated measurement schemes, as will be shown below.

# 5 Resolution of a spectrometer

In sect. 4.1, the simple measurement of the laser-light wavelength was used to illustrate the basic concepts of the Fisher information and its relations to statistical inference. In this section, we discuss the implications of the Fisher information in evaluating the performance of spectrometers, which are devices used to observe the spectrum of an incoming signal, as sketched in fig. 5. Owing to the fact that a spectrum carries unique information about the material from which light was emanated, spectroscopy is an essential tool for material analysis.

The spectral resolution is an important parameter that determines the spectrometer performance. This parameter quantifies the spectrometer's ability to distinguish between two closely spaced spectral bands and observe fine structures in the spectrum. Just like the case for two-point resolution, the standard spectral resolution criteria are based on the widths of the spectral bands at the output of the spectrometer and do not take data post-processing methods into account. In practice, to identify the chemical composition of a sample, its corresponding spectrum does not need to be fully resolved. Characteristic information about the sample can alternatively be retrieved by digitally matching its spectrum to a set of previously recorded spectra [15–18]. In this section, the Fisher information will be employed to discuss spectral resolution in detail.

To study the resolution of a spectrometer, it is necessary to understand how this device works. The key component of the spectrometer is an optical prism or diffraction grid—a dispersive element that spatially separates the individual spectral components of the measured signal. Other important components include the input slit, collimating and focusing optical elements and a position-sensitive photodetector. An incoming signal passes through the slit placed in the focal plane of a lens and is collimated before impinging on the dispersive element. Assuming a plane wave incident on a grating, the angle of deflection θ' for the first diffraction order is governed by the grid equation

$$\sin \theta' = f_g \lambda + \sin \theta \tag{21}$$

and depends on the signal wavelength λ, inclination angle θ and spatial frequency of the grid f_g. The output signal is collected by the second lens, whose optical axis is adjusted to coincide approximately with the direction of propagation
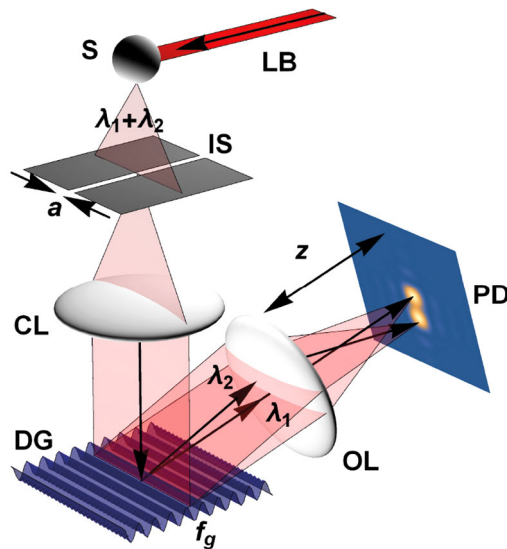
Fig. 5. Observation of the spectrum of incoming signals. The specimen (S), which is illuminated by the laser beam (LB), emanates signals consisting of two closely separated spectral bands. The signals pass through input slit (IS) and are then collimated with the collimating lens (CL) onto diffraction grid (DG). The resulting diffracted signals of the first diffraction order is collected by the objective lens (OL) and focused on the position-sensitive detector (PD), where single spectral components are spatially dispersed.

for the diffraction order $+1$, and focused on a photodetector. Thereafter, the spectral component $\lambda' = \lambda + \Delta\lambda$ is imaged at a position

$$x = \gamma z f_{\mathrm{g}} \Delta\lambda + \gamma z \sin\theta \tag{22}$$

with respect to the optical axis of the focusing lens. Here z is the focal length and $\gamma = 1/\sqrt{1 - (f_{\mathrm{g}}\lambda)^2}$ is a scaling factor that arises from the paraxial approximation for the sine function in eq. (21). This factor approaches unity for small f_{g}. According to eq. (22), the spectral bands separation is proportional to the frequency f_{g} of the grating and focal length z of the focusing element. One effect that needs to be considered during the discussion of spectrometer resolution is image blurring due to the different incident angles θ of the signal. Blurring arises from the θ-dependent term in eq. (21). In order to avoid this false line splitting, the input signal is collimated before it intersects the grating. The quality of collimation strongly depends on the width a of the collimator slit. Another important element of the imaging path is the excitation laser, which illuminates the specimen. Consequently, interaction of a laser beam with the specimen results in a set of spectral bands that encode the chemical signatures of the specimen. Imperfect temporal coherence is reflected in the broadening of the spectral bands of interest.

Finally, the image of the signal is obtained through convoluting the signal bands broadened by the excitation laser profile with the scaled aperture function of the entrance slit (again, the parameter γ plays the role of the scaling factor here):

$$I_{\mathrm{i}}(x) = \int \mathrm{d}x' I_{\mathrm{o}}(x') \frac{1}{a} \mathrm{rect}\left(\frac{x' - \Phi(x)}{a}\right), \qquad \Phi(x) = \frac{x}{\gamma} - \Delta\lambda z f_{\mathrm{g}}. \tag{23}$$

Assuming that the laser spectral band has a Gaussian profile of standard deviation b, the observed intensity turns out to be

$$I(x) = I_{\mathrm{i}}(x) \propto \mathrm{erf}\left(\frac{\Phi(x) + a/2}{\sqrt{2}z f_{\mathrm{g}} b}\right) - \mathrm{erf}\left(\frac{\Phi(x) - a/2}{\sqrt{2}z f_{\mathrm{g}} b}\right), \tag{24}$$

where $\mathrm{erf}(\cdot)$ is the error function. Both the collimating and the focusing components are assumed to be free of aberrations.

A good spectral resolution in the recorded data is not always necessary. Information about the sample is encoded not only in the number and position of individual spectral intensity peaks, but in the intensity profile as a whole. Not surprisingly, the Fisher information and the corresponding CRLB reveal this information. The discrete structure of the photodetector and the presence of noise in the data are incorporated in such an analysis. This is important because the detected signals are often very weak.
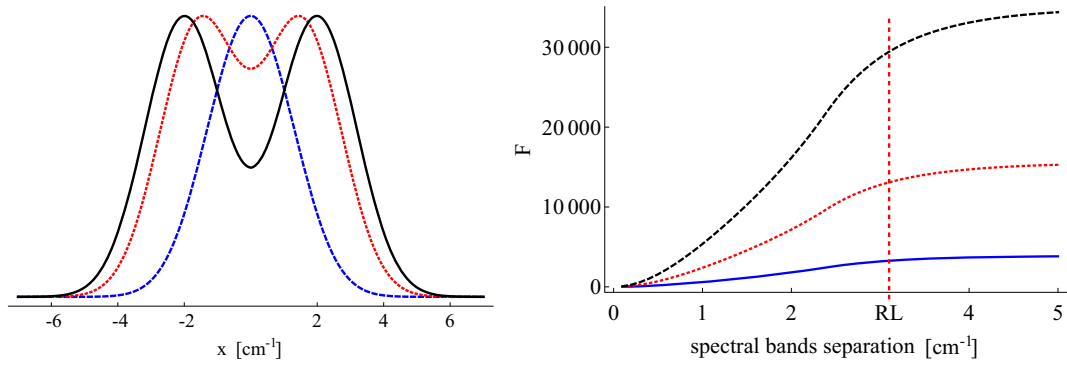
Fig. 6. Raman spectroscopy resolution. Left panel shows the output data for spectral doublets corresponding to three different separations: $1\,\mathrm{cm}^{-1}$ (dashed line), the Rayleigh separation $3.1\,\mathrm{cm}^{-1}$ (dotted line), and $4\,\mathrm{cm}^{-1}$ (solid line). For Gaussian peaks, the Rayleigh criterion that defines the Rayleigh separation is the attainment of a local minimum at 81% of the height of either peak. Right panel shows the Fisher information corresponding to estimating separations of the Raman spectral bands. Plots for three values of the noise-to-signal ratio NSR = 0.03 (solid line), 0.01 (dotted line), and 0.008 (dashed line) are presented. The parameters of the Raman spectrometer are $a = 5\,\mu\mathrm{m}$, $f_\mathrm{g} = 3600\,\mathrm{lines/mm}$, $z = 100\,\mathrm{mm}$, and $b = 1\,\mathrm{cm}^{-1}$. The camera pixel size is $4 \times 4\,\mu\mathrm{m}$ and the central pixel corresponds to a wavelength $\lambda = 266\,\mathrm{nm}$.
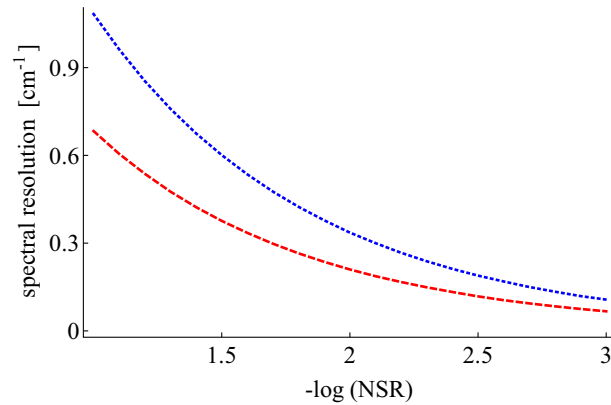
Fig. 7. The CRLB resolution limit as a function of the logarithmic noise-to-signal ratio (log(NSR)) for two different bandwidths of the excitation laser: $b = 2\,\mathrm{cm}^{-1}$ (dotted line) and $b = 1\,\mathrm{cm}^{-1}$ (dashed line). Notice that the limit strongly depends on the quality of the data.

The results of simulated spectroscopy of a pair of spectral bands is illustrated in fig. 6. The left panel shows the intensity profiles for three different separations of a pair of broadened spectral bands. The full-width at half maximum (FWHM) of each spectral band is $2.8\,\mathrm{cm}^{-1}$. This corresponds to a Rayleigh separation of approximately $3.1\,\mathrm{cm}^{-1}$. The separation of the two bands can be estimated from the data. The right panel shows the Fisher information as a function of the true separation for three different noise levels. Here, Poissonian detection statistics is assumed. Obviously, the information about d gets smaller with decreasing band separation and increasing noise level. However, some information about the separation is always extractable, even in regimes well beyond the standard Rayleigh limit.

To turn the Fisher information about the spectral-band separation into a resolution limit that can be directly compared with the standard Rayleigh limit, we consider a simple hypothesis testing method based on the CRLB. We denote by σ the CRLB for an efficient unbiased estimator for d. With reference to the CRLB, the spectral bands are resolved (unresolved) if the zero separation $d = 0$ falls outside inside the interval $[d - 3\sigma, d + 3\sigma]$ about the true value. This limit is evaluated in fig. 7 for two different excitation laser bandwidths. Comparing the blue solid line with the left panel of fig. 6, we emphasize that clever postprocessing of high-quality data brings about a resolution gain of at least an order of magnitude with respect to the standard heuristic resolution limits.

6 Superresolution of extended objects

Consider the imaging of an extended object of amplitude U_o by a linear spatially-invariant optical system described by the impulse response function $h(\cdot)$. The input-output transformation in eq. (1) for intensities (incoherent imaging), or complex amplitudes (coherent imaging), is considerably simplified in the spatial frequency domain,

$$G_\mathrm{i}(f_x, f_y) = H(f_x, f_y)G_\mathrm{o}(f_x, f_y). \tag{25}$$
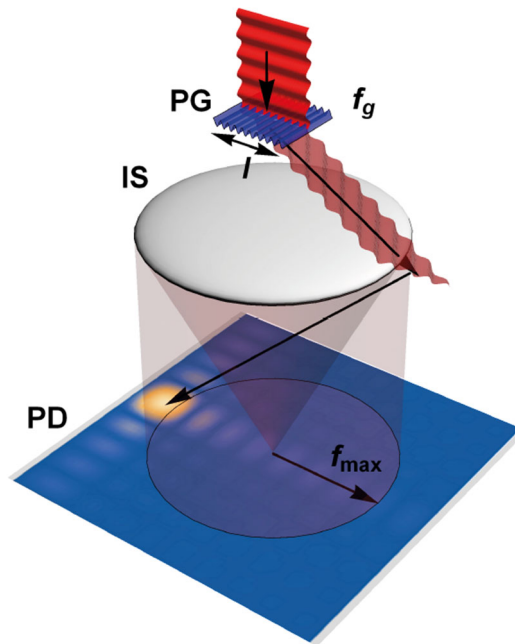
Fig. 8. Superresolution of the optical imaging system (IS). The spatially bounded phase grating (PG) is illuminated by a plane wave. If the grid is blazed, only signals of one diffraction order are propagated toward the imaging system. Even if the frequency of the phase grating is higher than the cutoff frequency of the imaging system, information about the frequency is still present in the position-sensitive detector (PD), which is placed in the focal plane of the imaging system.

Physically, the spectrum of the image G_i is obtained by simply applying a frequency filter, described by a transfer function $H(\cdot)$, to the spectrum of the object G_o. Transfer functions represent important characteristics of imaging systems. They describe how well the individual spatial frequencies of the object spectrum are mapped to the final image spectrum, and are thus related to the spectral resolution which was discussed previously. Indeed, fine details, sharp edges, etc. appearing in the object spectrum correspond to high-frequency components. The filtering of these components thus results in the blurring of the object spectrum and a reduction in resolution. The standard resolution criterion is based on the highest frequency transmitted f_{max}.

Without loosing generality, we focus on spatially invariant diffraction-limited coherent imaging systems, whose transfer functions can be, under mild conditions, shown to be scaled versions of the system aperture functions. Generalization to incoherent illumination or imperfect systems is straightforward. The finite spatial extent of any real aperture means that the corresponding transfer function is bounded and there is always a frequency cutoff beyond which spectral transfer drops to zero. Signals of higher frequencies are completely filtered out from the image and the imaging system therefore acts as a low-pass frequency filter. One would think that the resolution of a real imaging system is always limited to the size of the cutoff frequency period $1/f_{max}$ and an object modulation faster than this cutoff is not observable. Fortunately, it turns out that the resolution can be extended far beyond this limit, provided that some prior information about the object is available.

To keep the analysis as simple as possible we discuss imaging of a simple 1-D periodic object modulation $U_o(x) = \exp(-i2\pi f_g x)$, where f_g is the modulation frequency (see fig. 8). This frequency is assumed to exceed the cutoff frequency f_{max} of the imaging system with a finite rectangular aperture ($f_g > f_{max}$). Since the corresponding object spectrum $G_o(f) \propto \delta(f - f_g)$ is a delta function peaked at f_g, no information about the object is available in the image:

$$G_i(f) = \text{rect}\left(\frac{f}{2f_{max}}\right)\delta(f - f_g). \tag{26}$$

The situation drastically changes with a realistic object of finite size. In this case, the delta-function spectrum of the oscillatory modulation is broadened by a convolution with the rectangular function defining the object length ℓ: $U_o(x) \mapsto \text{rect}(x/\ell)U_o(x)$ and $G_o(f) \propto \text{sinc}[\ell(f - f_g)]$. Although the spectrum remains centered outside the support of $H(f)$, notice that the "tails" of the sinc function overlap with the transmitted range of frequencies (see fig. 9) and information about f_g is now accessible despite the fact that f_g is larger than the cutoff f_{max}, since

$$G_i(f) = \text{rect}\left(\frac{f}{2f_{max}}\right)\text{sinc}[\ell(f - f_g)]. \tag{27}$$
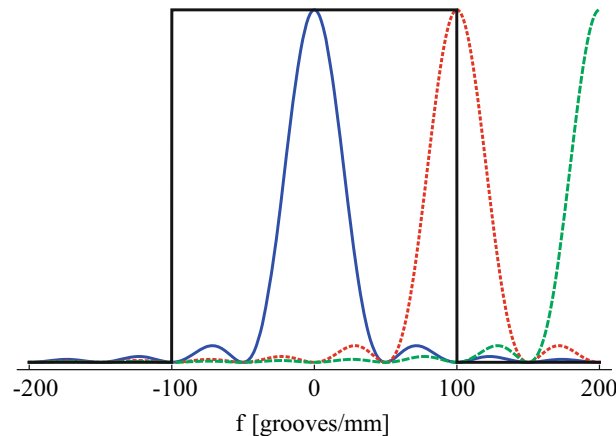
Fig. 9. Spatial spectra of three different amplitude gratings. The transfer function H of the imaging system is shown with the black rectangle. Signals modulated in frequencies inside the transmitted range can be observed.
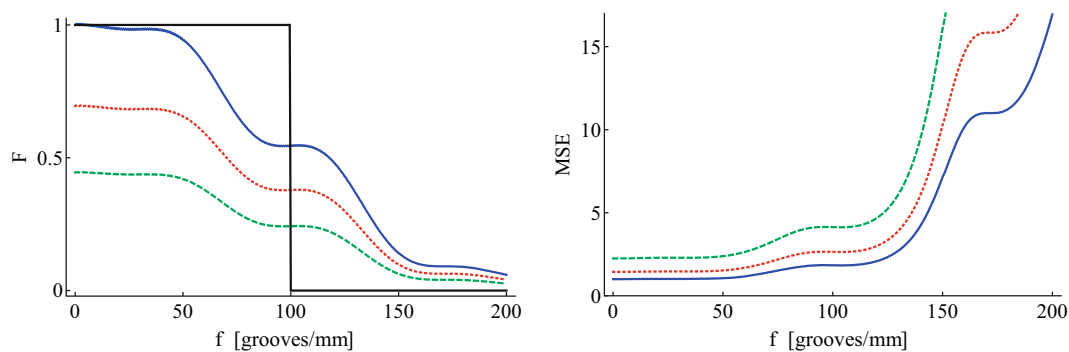

Fig. 10. Fisher information and errors associated with estimating the grid modulation frequency. The transfer function and the grating size are the same as in fig. 9. In total, 1000 equidistant samples of the image spectral intensities are processed for each grating. Poissonian statistics is considered. All values are normalized to the first estimation (solid line) corresponding to constant grid amplitude (true frequency $f = 0$) and to the particular NSR. Plots for 1.2NSR (dotted line), and 1.5NSR (dashed line) are also presented.

This discussion can be generalized to any spatially bounded object [9]. The Fourier spectrum of any bounded function is unbounded and analytic. As a consequence, the full spectrum can, in principle, be reconstructed by analytic continuation from the observable frequency region. The Fisher information can be used to quantify the resolution limits of such superresolution techniques.

To demonstrate this, let us return to the example of an amplitude grating of a finite size ℓ. This example is sketched in fig. 9. Only diffraction of a single order shall concern us here, which is a consequence of illuminating a blazed grid with a plane wave. The present goal is to estimate the grid frequency f_g from the image spectrum recorded with a position-sensitive photodetector. In the superresolution regime, most object intensity is filtered out according to the system transfer function and noise becomes an important factor. The results that are summarized in fig. 10 show the normalized Fisher information and the MSE against the true frequency and the data noise-to-signal ratio NSR. By increasing the modulation frequency f_g of the object, the Fisher information decreases and the MSEs become larger, as expected. The higher the grid frequency f_g, the more difficult it is for its estimation. In addition, we observe that spectral observations in the standard resolution regime and the superresolution regime differ strongly as far as the sensitivity of the measurement to data noise is concerned: While increasing the noise from NSR = 100 to NSR = 10 has little effect on the estimation of f_g at 50% of the cutoff frequency f_{max}, for modulations at 130% of f_{max}, the MSE is more than doubled and the inference of even faster modulations require data that are almost noiseless. As such, we conclude that resolution limits of a given imaging system cannot be determined before the noise properties and a statistical model for the signal detection are specified, contrary to the overly simplified treatments based on the geometry of the PSF alone that can be found in some standard textbooks (see [3] for instance).

Finally, in fig. 11, the role of the object size in superresolution estimation is discussed. For objects of infinite size, the Fisher information about the frequencies exceeding the cutoff is zero. By decreasing the size of the object, information about low-frequency features is reduced, yet inference in the superresolution regime becomes possible. Smaller object sizes, or large amount of prior information about the object, allow for extending the resolution far beyond the imaging-system cutoff. As mentioned above, high-quality data with very little noise and other systematic errors are required to exploit this benefit.
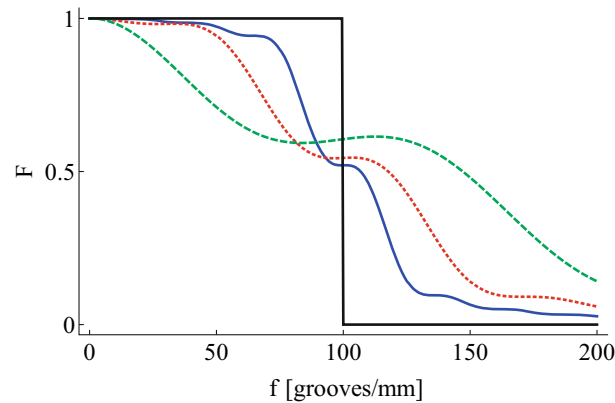
Fig. 11. Estimation of the grid modulation frequency. The Fisher information is shown for grids of different sizes corresponding to different FWHM spectral widths: 50 grooves/mm (solid line), 100 grooves/mm (dotted line), and 200 grooves/mm (dashed line). The cutoff frequency is the same as in figs. 9 and 10. The spectrum was sampled at 1000 equidistant frequencies and the calculated Fisher information values are normalized to unity at the true frequency $f = 0$.

7 Biased estimation strategies: beating the Cramér-Rao Lower Bound

Thus far, this work has been devoted to the discussion of the CRLB and its implications in the analysis of optical measurements. These results are, strictly speaking, only applicable to unbiased estimation strategies. To complete picture, we shall remind the reader of the possibility of beating the statistical inference limits set by the CRLBs by looking at estimators that are *biased*.

In sect. 4.1, we investigated the inference limit of wavelength estimation by considering angular measurements of two diffraction spots of orders $m = 1$ and $m = -1$. This limit is defined by the CRLB in eq. (16). Evidently, the linear unbiased estimator that saturates this inequality is given by $\widehat{\lambda} = d(\phi_1 - \phi_{-1})/2$, since

$$(\Delta\widehat{\lambda})^2 = \frac{d^2}{2}(\Delta\phi)^2. \tag{28}$$

This efficient estimator is also known as the best linear unbiased estimator (BLUE) [19,20] that minimizes the variance for the problem. One can understand why this linear estimator is the BLUE by recognizing that this estimator maximizes the *likelihood* $p(\phi_1, \phi_{-1}|\lambda)$, which is in fact a Gaussian distribution for the random vector $(\phi_1, -\phi_{-1})$ with mean vector $(\lambda/d, \lambda/d)$ and covariance matrix $\mathrm{diag}(\sigma^2, \sigma^2)$, and so the statement of asymptotic efficiency, which states that the maximum-likelihood (ML) estimator of a Gaussian likelihood achieves the CRLB, holds [21].

On the other hand, one may choose another estimator

$$\widehat{\lambda} = d\sqrt{\frac{\phi_1^2 + \phi_{-1}^2}{2}}, \tag{29}$$

which takes the root-mean-square (RMS) of the data. This estimator is no longer unbiased since $\overline{\overline{\lambda}} \neq \lambda$. One can check that the variance of this positive estimator is given by

$$\left(\frac{\Delta\widehat{\lambda}}{d}\right)^2 = \sigma^2 + 2\mu^2 - \sqrt{\pi}\frac{\mu}{\sigma}e^{-\frac{\mu^2}{2\sigma^2}}\left[(\sigma^2 + \mu^2)\,\mathrm{I}_0\left(\frac{\mu^2}{2\sigma^2}\right) + \mu^2\mathrm{I}_1\left(\frac{\mu^2}{2\sigma^2}\right)\right] \tag{30}$$

with $\mu = \lambda/d$ and $\mathrm{I}_n(y)$ is the modified Bessel function of the first kind of order n. Note that with this positive estimator, it is possible to achieve an estimation precision smaller than the CRLB. For simplicity, let us take the spacing d to be of unit length and λ to be a scaled wavelength relative to d. When $\lambda \gtrsim \Delta\phi/2$, the MSE for the biased estimator in eq. (29) is always smaller than that for the BLUE, with the largest improvement typically near $\lambda = \Delta\phi$ and $(\Delta\widehat{\lambda})^2 \lesssim (\Delta\phi)^2/2 = \mathrm{CRLB}$ in the "infrared" regime ($\lambda \gg \Delta\phi$). More generally, under certain conditions, it is possible to find a more accurate biased estimator that gives a lower MSE than the CRLB for unbiased estimators. This possibility is not surprising and is well-known in the theory of statistical inference for decades [22–24].
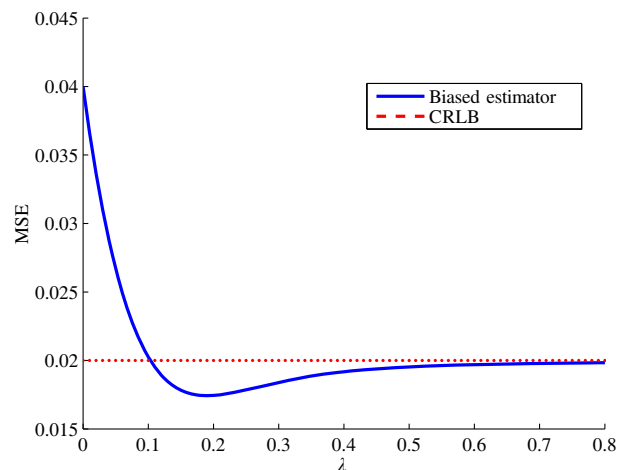
Fig. 12. Plots of the variance of the biased estimator in eq. (29) and the CRLB that is achieved by the BLUE against the scaled wavelength λ for $\Delta\phi = 0.2$. For $\lambda > 0.103 \approx \Delta\phi/2$, the biased estimator in eq. (29) is more accurate than the BLUE, with the lowest MSE of 0.0174 occurring at $\lambda = 0.190 \approx \Delta\phi$. The physics of the diffraction problem applies more accurately to the region $\lambda \gtrsim \Delta\phi$.

When the spots of the two diffraction orders are closely spaced ($\lambda \lesssim \Delta\phi/2$), the MSE for the biased estimator is higher than that of the BLUE and we have $(\Delta\widehat{\lambda})^2 \lesssim (\Delta\phi)^2 = 2 \times \text{CRLB}$ in the "ultraviolet" regime ($\lambda \ll \Delta\phi$), so that the BLUE is the better choice (see fig. 12). However, note that the results of eqs. (16) apply to the condition that the scaled λ is much larger than the angular uncertainty $\Delta\phi$. In a more realistic experiment, when the orders are spaced too closely, the diffraction spots overlap and are in general impossible to be distinguished, so that the angular measurement of a specific diffraction order loses its meaning. Therefore, these reasonings make physical sense only for moderate to large values of the scaled λ relative to $\Delta\phi$, for which a biased estimation strategy may prove to be much more reliable for the same data.

As an additional assurance, we mention here that if the objective is to supply an estimator that gives the minimal MSE possible, then the bias of this estimator is not a real problem as long as it is a consistent estimator. For the biased estimator in eq. (29), as long as the accuracy of the measured values ϕ_1 and ϕ_{-1} are high, this estimator approaches the true value.

8 Conclusions

We have discussed, in this tutorial review, various applications of the Fisher information in optics. Through these applications, we highlighted that the Cramér-Rao lower bound that is defined by this information measure defines the limits of optical resolution, both for the standard Rayleigh limit and the superresolution limit. These concepts also give rise to fundamental inequalities that appear in wave optics and are particularly useful in evaluating the performance of optical imaging systems. As a final note, we remind the reader that more precise estimation strategies can be employed to beat the limit set by the Cramér-Rao bound in statistical inference if some prior information about the object of interest is known, and in some cases, there is remarkably large improvement for the corresponding estimators over those obtained with common unbiased estimation strategies.

References

1. "Information" `Merriam-Webster.com` (Merriam-Webster, 2015) Web. 6 November 2015.
2. "Accuracy" `Merriam-Webster.com` (Merriam-Webster, 2015) Web. 6 November 2015.
3. M. Born, E. Wolf, *Principles of Optics* (Cambridge, Cambridge University Press, 1999).
4. A.J. den Dekker, A. van den Bos, J. Opt. Soc. Am. A **14**, 547 (1997).
5. C.R. Rao, Bull. Calcutta Mat. Soc. **37**, 81 (1945).
6. H. Cramér, *Mathematical Methods of Statistics* (Princeton, Princeton University Press, 1946).
7. B.R. Frieden, *Probability, Statistical Optics, and Data Testing* (Springer, Berlin, 1983).
8. P. Facchi, R. Kulkarni, V.I. Manko, G. Marmo, E.C.G. Sudarshan, F. Ventriglia, Phys. Lett. A **374**, 4801 (2010).
9. J.W. Goodman, *Introduction to Fourier Optics* (Roberts & Company, Englewood, 2004).
10. J. Řeháček, Z. Hradil, J. Mod. Opt. **51**, 979 (2004).
11. W. Heisenberg, *The Physical Principles of the Quantum Theory*, translated by C. Eckart, F.C. Hoyt (Dover, New York, 1949).
12. N. Bohr, Nature **121**, 580 (1928).
13. B.-G. Englert, *Lectures on Quantum Mechanics—Volume 3: Perturbed Evolution* (World Scientific, Singapore, 2006).
14. T. Durt, B.-G. Englert, I. Bengtsson, K. Życzkowski, Int. J. Quantum Inf. **8**, 535 (2010).
15. J. Gordon Robertson, Publ. Astron. Soc. Aust. **30**, e048 (2013).
16. M. Fuller, R. Rosenthal, Proc. SPIE **2089**, 440 (1993).
17. W. Fremout, S. Saverwyns, J. Raman Spectrosc. **43**, 1536 (2012).
18. J. Hodkiewicz, F. Deck, *Evaluating Spectral Resolution on a Raman Spectrometer* (Thermo Fisher Scientific, Madison, 2010).
19. S. Puntanena, G.P.H. Styan, Amer. Stat. **43**, 153 (1989).
20. C.R. Rao, J. Multivar. Anal. **3**, 276 (1973).
21. R. Fisher, Phil. Trans. R. Soc. A **222**, 309 (1921).
22. A.E. Hoerla, R.W. Kennarda, Technometrics **12**, 55 (1970).
23. R.F. Gunst, R.L. Mason, J. Am. Stat. Assoc. **72**, 616 (1977).
24. S. Kay, Y.C. Eldar, IEEE Signal Process. Mag. **25**, 133 (2008).