

Estimating random signal parameters from noisy images with nuisance parameters: linear and scanning-linear methods

Meredith Kathryn Whitaker, Eric Clarkson,
and Harrison H. Barrett

College of Optical Sciences and Department of Radiology, University of Arizona,
Tucson, Arizona 85724, USA

mwhitaker@optics.arizona.edu

Abstract: In a pure estimation task, an object of interest is known to be present, and we wish to determine numerical values for parameters that describe the object. This paper compares the theoretical framework, implementation method, and performance of two estimation procedures. We examined the performance of these estimators for tasks such as estimating signal location, signal volume, signal amplitude, or any combination of these parameters. The signal is embedded in a random background to simulate the effect of nuisance parameters. First, we explore the classical Wiener estimator, which operates linearly on the data and minimizes the ensemble mean-squared error. The results of our performance tests indicate that the Wiener estimator can estimate amplitude and shape once a signal has been located, but is fundamentally unable to locate a signal regardless of the quality of the image. Given these new results on the fundamental limitations of Wiener estimation, we extend our methods to include more complex data processing. We introduce and evaluate a scanning-linear estimator that performs impressively for location estimation. The scanning action of the estimator refers to seeking a solution that maximizes a linear metric, thereby requiring a global-extremum search. The linear metric to be optimized can be derived as a special case of maximum *a posteriori* (MAP) estimation when the likelihood is Gaussian and a slowly varying covariance approximation is made.

© 2008 Optical Society of America

OCIS codes: (110.3000) Image-Quality Assessment; (100.2960) Image Analysis; (110.4280) Noise in Imaging Systems

References and links

1. H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley-Interscience, 2004).
2. S. C. Moore, M. F. Kijewski, and G. El Fakhri, "Collimator Optimization for Detection and Quantitation Tasks: Application to Gallium-67 Imaging," *IEEE Trans. Med. Imaging* **24**(10), 1347–1356 (2005).
3. G. El Fakhri, S. C. Moore, and M. F. Kijewski, "Optimization of Ga-67 imaging for detection and estimation tasks: Dependence of imaging performance on spectral acquisition parameters," *Med. Phys.* **29**, 1859–1866 (2002).
4. M. A. Kupinski, E. Clarkson, K. Gross, and J. W. Hoppin, "Optimizing imaging hardware for estimation tasks," *Proc. SPIE* **5034**, 309–313 (2003).

5. H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," *J. Opt. Soc. Am. A* **7**(7), 1266–1278 (1990).
6. H. H. Barrett, K. J. Myers, N. Devaney, and J. C. Dainty, "Objective Assessment of Image Quality: IV. Application to Adaptive Optics," *J. Opt. Soc. Am. A* **23**(12), 3080–3105 (2006).
7. Y. C. Eldar, "Comparing between estimation approaches: admissible and dominating linear estimators," *IEEE Trans. Signal Process.* **54**(5), 1689–1702 (2006).
8. C. Caiafa, A. Proto, and E. Kuruoglu, "Long correlation Gaussian random fields: Parameter estimation and noise reduction," *Digital Signal Processing* **17**, 819–835 (2007).
9. A. B. Hamza, H. Krim, and G. Unal, "Unifying probabilistic and variational estimation," *IEEE Signal Process. Mag.* **19**, 37–47 (2002).
10. R. E. Greenblatt, A. Ossadtchi, and M. E. Pflieger, "Local linear estimators for the bioelectromagnetic inverse problem," *IEEE Trans. Signal Process.* **53**(9), 3403–3412 (2005).
11. F. O. Bochud, J.-F. Valley, and F. R. Verdun, "Estimation of the noisy component of anatomical backgrounds," *Med. Phys.* **26**(7), 1365–1370 (1999).
12. S. P. Müller, M. F. Kijewski, S. C. Moore, and B. L. Holman, "Maximum-likelihood Estimation: A Mathematical Model for Quantitation in Nuclear Medicine," *J. Nucl. Med.* **31**, 1693–1701 (1989).
13. S. P. Müller, C. K. Abbey, F. J. Rybicki, S. C. Moore, and M. F. Kijewski, "Measures of performance in nonlinear estimation tasks: prediction of estimation performance at low signal-to-noise ratio," *Phys. Med. Biol.* **50**, 3697–3715 (2005).
14. J. Shao, *Mathematical Statistics* (Springer, 1999).
15. M. Kupinski, J. W. Hopkin, E. Clarkson, and H. H. Barrett, "Ideal-observer computation in medical imaging with use of Markov-chain Monte Carlo techniques," *J. Opt. Soc. Am. A* **20**, 430–438 (2003).
16. N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications* (The MIT press, 1949).
17. J. L. Melsa and D. L. Cohn, *Decision and Estimation Theory* (McGraw-Hill, 1978).
18. J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," *J. Opt. Soc. Am. A* **9**, 649–658 (1992).
19. H. H. Barrett, K. J. Myers, B. Gallas, E. Clarkson, and H. Zhang, "Megalopiniakophobia: Its Symptoms and Cures," *Proc. SPIE* **4320**, 299–307 (2001).

1. Introduction

The evaluation of image quality is rigorously approached by first stipulating the scientific purpose the images are meant to serve. We will consider estimation tasks, *i.e.*, use of the image data to estimate features of the object. Once the imaging system produces measurements of the object, the estimates are derived from the image by an observer. Many types of observers are used for estimation tasks: a person selecting a region-of-interest from a data display, an automated computer algorithm, or some hybrid of the two. A task-based measure of image quality analyzes the ability of an observer to perform a specified task [1]. In this paper, we will analyze two mathematical observers (Wiener and scanning linear) and use the ensemble mean squared error (EMSE) as a figure of merit to quantify observer performance. We will show that these estimators minimize the EMSE under different statistical assumptions and offer potential methods for image-quality evaluation.

Estimating parameters of interest from image data encompasses many fields of imaging. In medical imaging, images are diagnostic tools used for tasks such as cancer detection, assessment of disease state, and treatment efficacy. These tasks often require quantification of diagnostically relevant parameters such as tumor volume, tracer uptake in a region of interest, and tumor location. Estimation tasks in astronomy include using image data to estimate the photometric output of celestial objects. Satellite data are used to estimate features of the Earth's landscape or to locate targets. These estimates are commonly calculated using simple mathematics on a reconstructed image, such as summing the pixel values within a user-specified region or selecting a maximum pixel value. Such estimation procedures were invented to produce numbers quickly and easily, but are not derived according to cost minimization and are usually not evaluated using quantitative measures of performance. Quantitative assessment of estimation performance has been used to make design decisions for imaging hardware [2], [3],

[4]. More specifically, the Wiener estimator has been suggested for image-quality assessment [5], [6], [7], [8] but this paper will show its limitations. Motivated by the cases when linear Wiener estimation fails, we have developed an estimation rule that seeks a global maximum to a linear operation on the data. The search for this extremum is alluded to by calling this new estimator *scanning linear*.

Measurement noise always influences the statistics of the image and thus all subsequent estimates derived from noisy images. In recent years, a variety of linear and nonlinear techniques have been proposed to tackle estimation problems in a noisy image environment; see for example [9], [10]. To complicate the task further, for a majority of realistic imaging problems, even the noise-free data will not be uniquely determined by the parameters to be estimated. The image is usually influenced by effects that interfere with pure measurements of the desired quantities. Such effects are called nuisance parameters and can include features like non-targeted uptake in nuclear medical imaging, atmospheric turbulence in astronomical applications, or any attribute of the object that is not being estimated yet influences the image data in a manner that is statistical and cannot be accounted for as a simple correction factor. In our current study, the object is composed of a signal and a background term. The background represents the nuisance parameters because it conveys no information about the signal parameters we wish to estimate, yet it is a source of image variability and therefore affects the statistics of our estimate [11]. The data likelihood is effected by nuisance parameters, and this greatly complicates rigorous maximum-likelihood (ML) estimation [13]. Calculating the data likelihood requires averaging over all the nuisance parameters, and furthermore this average must be recalculated to produce a candidate ML estimate from each image realization. Multiple candidate solutions are generated to search for the one which maximizes the likelihood. The scanning linear estimator introduced in this paper relies on a Gaussian likelihood assumption to avoid the marginalization over nuisance parameters.

We begin in Section 2 by deriving the Wiener estimator and the scanning linear estimator as solutions to an EMSE minimization. The imaging equation, a description of measurement noise in imaging systems, and the object model are presented in Section 3. The object is decomposed into a signal and a background term. The signal is parameterized by the object features we wish to estimate: the signal size, amplitude, and location. The background represents nuisance parameters. In addition to measurement noise, the signal and the background are also random. This leads to a triply stochastic treatment of the image data, which is described in Section 4. This section also defines the first- and second- order statistics of the data as well as a cross-covariance between the images and the parameters to be estimated. At this juncture we transition from theoretical forms of the estimation rules to the specifics of our simulation study. The models and statistics of the simulated objects are introduced in Section 5. Section 6 provides the details of our simulated imaging operator with a display of sample images, data covariances, cross-covariance, and the model-specific calculation of the scanning linear estimator. Finally, the performance of both estimators, for various cases of prior knowledge, is illustrated and discussed in Section 7. The conclusions are presented in Section 8.

2. Deriving the estimation rules

We will consider a digital image, written as an $M \times 1$ vector \mathbf{g} , that has some general relationship to a set of parameters denoted by an $N \times 1$ vector called $\boldsymbol{\theta}$, with elements corresponding to the parameters to be estimated. An estimate of these parameters, calculated from an image, will be denoted by $\hat{\boldsymbol{\theta}}(\mathbf{g})$. The vector $\boldsymbol{\theta}$ will be treated as a random variable drawn from an ensemble described by a probability density function (pdf) called $\text{pr}(\boldsymbol{\theta})$. The performance of the estimation scheme will be evaluated according to its average performance on the population, rather than for one specific object.

Throughout this paper, statistical averages will be written as

$$\langle y(\mathbf{x}) \rangle_{\mathbf{x}} = \int d^P \mathbf{x} \text{pr}(\mathbf{x}) y(\mathbf{x}), \quad (1)$$

where \mathbf{x} is an $P \times 1$ random vector with pdf $\text{pr}(\mathbf{x})$, $y(\mathbf{x})$ is a scalar-valued function of the random vector, and the subscript on the LHS conveys the ensemble used for the expectation. This notation allows for expectations with respect to multiple random quantities to be expanded into nested conditional averages in a short-hand form, such as $\langle \cdot \rangle_{\mathbf{g}, \theta} = \langle \langle \cdot \rangle_{\mathbf{g}|\theta} \rangle_{\theta}$.

The performance of an estimation rule is conventionally quantified by measuring the distance between the estimate and the true value of the parameter. The mean-squared error (MSE) is a popular choice for quantifying this distance. The stochastic nature of this problem is accounted for by calculating the EMSE,

$$\text{EMSE}(\hat{\theta}(\mathbf{g})) = \left\langle \left\langle \|\hat{\theta}(\mathbf{g}) - \theta\|^2 \right\rangle_{\mathbf{g}|\theta} \right\rangle_{\theta} = \text{tr} \left\langle \left\langle (\hat{\theta}(\mathbf{g}) - \theta) (\hat{\theta}(\mathbf{g}) - \theta)^t \right\rangle_{\mathbf{g}|\theta} \right\rangle_{\theta}. \quad (2)$$

We have made use of the equality between an inner product of two vectors and the trace of their outer product. The trace, defined as the sum of the diagonal elements of a square matrix, is denoted tr . The EMSE calculation involves the familiar squared difference between estimated and actual values; however because the estimation task can in general be multi-dimensional, the measure of difference is a vector norm.

Among all possible estimators, the mean of the posterior density $\text{pr}(\theta|\mathbf{g})$ minimizes EMSE and is given by

$$\hat{\theta}(\mathbf{g})_{\text{PM}} = \underset{\hat{\theta}(\mathbf{g})}{\text{argmin}} \left[\text{EMSE}(\hat{\theta}(\mathbf{g})) \right] = \int d^N \theta \text{pr}(\theta|\mathbf{g}) \theta, \quad (3)$$

where the subscript PM denotes the posterior mean, and N is the number of parameters to be estimated. This estimation rule is, in general, nonlinear with respect to \mathbf{g} and requires sufficient knowledge of the posterior density to calculate or estimate the integral expression. Markov-chain Monte Carlo (MCMC) techniques can be used to estimate the posterior mean [14], [15], but the disadvantages include computational time and stringent requirements on prior knowledge.

In this paper, we will evaluate two estimation rules that minimize EMSE under different statistical conditions that simplify the form of the posterior mean.

2.1. Wiener estimation

When the joint pdf of the image data and the parameters, denoted $\text{pr}(\theta, \mathbf{g})$, is Gaussian the posterior mean of Eq. (3) reduces to the linear Wiener estimator, given by [16],[17]

$$\hat{\theta}_W(\mathbf{g}) = \bar{\theta} + K_{\theta, \mathbf{g}} K_{\mathbf{g}}^{-1} (\mathbf{g} - \bar{\mathbf{g}}), \quad (4)$$

where the double overbar indicates averages over \mathbf{g} and θ as in Eq. (2). The optimal EMSE is given by

$$\text{EMSE}_W = \text{tr}[K_{\theta}] - \text{tr}[K_{\theta, \mathbf{g}} K_{\mathbf{g}}^{-1} K_{\theta, \mathbf{g}}^t]. \quad (5)$$

Here $K_{\theta, \mathbf{g}} = \langle (\theta - \bar{\theta})(\mathbf{g} - \bar{\mathbf{g}})^t \rangle_{\mathbf{g}, \theta}$ is the cross-covariance between the parameters and the data, which is a measure of the interdependence between these two random vectors. The cross-covariance is an $N \times M$ matrix, where N is the number of parameters to be estimated, and M is the number of elements in an image. In general, the cross-covariance is non-square because the number of parameters to be estimated is usually much less than the dimensionality

of an image. A given row of $K_{\theta, \mathbf{g}}$ selects one particular element of θ and is a measure of how changes to this scalar quantity will change each of the M elements in the image data. The cross-covariance is an expectation with respect to the data as well as the parameter ensemble; hence the calculation will require knowledge of the pdf $\text{pr}(\theta)$. Calculating the Wiener estimator also requires knowledge of the first- and second-order statistics of the data \mathbf{g} ; the inverse of the data covariance $K_{\mathbf{g}}^{-1}$ and the average data $\bar{\mathbf{g}}$ appear in the estimation rule of Eq. (4).

Among all affine transforms of the data (defined as a linear plus a constant term), the Wiener estimator minimizes the EMSE and, if the joint pdf is Gaussian, it minimizes the EMSE among all estimators. An analogy can be seen with detection tasks; the ideal observer requires full knowledge of the likelihood ratios, but restrictions to optimal linear test statistics reduce the requirement on prior knowledge to first- and second-order data statistics [5]. In both detection and estimation tasks, the optimal nonlinear performers, the ideal observer and posterior mean, are equivalent to their respective linear reductions, the Hotelling observer and Wiener estimator, when statistics are Gaussian [5].

2.2. Scanning-linear estimation

If the posterior density $\text{pr}(\theta|\mathbf{g})$ is unimodal and sufficiently symmetric about the mean, then the mean and mode of the distribution will be close. Calculating the mode of the posterior density (also called maximum *a posteriori* (MAP) estimation) requires a scan of parameter space to compare solutions and find the maximum. The general form of MAP estimation is

$$\begin{aligned}\hat{\theta}_{\text{MAP}} &= \underset{\theta}{\text{argmax}} \{ \text{pr}(\theta|\mathbf{g}) \} \\ &= \underset{\theta}{\text{argmax}} \left\{ \frac{\text{pr}(\mathbf{g}|\theta) \text{pr}(\theta)}{\text{pr}(\mathbf{g})} \right\} \\ &= \underset{\theta}{\text{argmax}} \{ \ln \text{pr}(\mathbf{g}|\theta) + \ln \text{pr}(\theta) \},\end{aligned}\quad (6)$$

where the second equality comes from using Bayes' rule and the third because $\ln(\cdot)$, the natural logarithm, is a monotonic transform and $\text{pr}(\mathbf{g})$ is independent of θ . The likelihood of the data, conditioned on the parameters, is called $\text{pr}(\mathbf{g}|\theta)$. In order to simplify the function to be optimized, we will approximate the likelihood by a Gaussian, but this condition does not imply that the joint pdf $\text{pr}(\mathbf{g}, \theta)$ is also Gaussian. The approximate likelihood is

$$\text{pr}(\mathbf{g}|\theta) \approx \frac{1}{\sqrt{(2\pi)^M \det(K_{\mathbf{g}|\theta})}} \exp \left[-\frac{1}{2} (\mathbf{g} - \bar{\mathbf{g}}(\theta))^t K_{\mathbf{g}|\theta}^{-1} (\mathbf{g} - \bar{\mathbf{g}}(\theta)) \right], \quad (7)$$

where $\bar{\mathbf{g}}(\theta)$ is the data averaged over all sources of randomness except the parameters to be estimated θ , $K_{\mathbf{g}|\theta}$ is the covariance matrix for the random vector $\mathbf{g} - \bar{\mathbf{g}}(\theta)$, and $\det(\cdot)$ is the determinant of the matrix. When searching for the maximum of this Gaussian likelihood we would have to re-evaluate the covariance, its inverse, and its determinant at every candidate θ . Our second approximation is to average over this dependence. This is equivalent to making a slowly varying parameter approximation

$$K_{\mathbf{g}|\theta} \approx \langle K_{\mathbf{g}|\theta} \rangle_{\theta} = \bar{K}_{\mathbf{g}}. \quad (8)$$

The determinant and other scale factors in Eq. (7) are now independent of θ and can be ignored for the purposes of an extremum search. In order to avoid the effort of exponentiating, we will operate on the natural logarithm of this approximate posterior density, which yields

$$\ln [\text{pr}(\theta|\mathbf{g})] \approx -\frac{1}{2} (\mathbf{g} - \bar{\mathbf{g}}(\theta))^t \bar{K}_{\mathbf{g}}^{-1} (\mathbf{g} - \bar{\mathbf{g}}(\theta)) + \ln[\text{pr}(\theta)] + \text{const.} \quad (9)$$

After dropping terms that do not depend on θ , the scanning-linear estimator which maximizes the approximate posterior density is

$$\hat{\theta}_{SL}(\mathbf{g}) = \operatorname{argmax}_{\theta} \left\{ \bar{\mathbf{g}}(\theta)^t \bar{\mathbf{K}}_{\mathbf{g}}^{-1} \mathbf{g} - \frac{1}{2} \bar{\mathbf{g}}(\theta)^t \bar{\mathbf{K}}_{\mathbf{g}}^{-1} \bar{\mathbf{g}}(\theta) + \ln [\operatorname{pr}(\theta)] \right\}. \quad (10)$$

The first term is a linear operation on the image data, given by a vector inner product with $\bar{\mathbf{g}}(\theta)^t \bar{\mathbf{K}}_{\mathbf{g}}^{-1}$. The next two terms are independent of the data, but must be retained because of the dependence on θ . The scanning-linear estimator is an increase in data-processing complexity compared to the Wiener estimator, or any other pure linear estimator. This estimator operates linearly on the data, though the linear template itself is a general nonlinear function of the parameters. The estimation rule requires a search for the value of θ that will maximize this linear operation on the data. Once the data statistics and parameterized signal model for our present study have been introduced, the specific form of this estimator will be presented in Section 6.3.

3. Assumptions and models

3.1. The imaging operator

An object can be represented as a continuous function of space and written as $f(\mathbf{r})$. The associated image is a finite number of measurements that may be organized into the vector \mathbf{g} . This leads to the mathematical description of imaging as an operator that maps a function of continuous variables to a discrete vector [1]. If the relationship between the spatial distribution of the object and the noise-free image data is also linear, this mapping can be written as [1]

$$[\bar{\mathbf{g}}]_m = \int f(\mathbf{r}) h_m(\mathbf{r}) d^q r, \quad (11)$$

where \mathbf{r} is a vector in q dimensions (*e.g.*, $q = 2$ or 3), and m is the index for the m^{th} measurement of the imaging system. This index can distinguish the various elements across a detector face. It can also be organized, in a lexicographical fashion, to include multiple cameras of tomographic systems or successive exposures of a temporal sequence of images. The function $h_m(\mathbf{r})$ is the system sensitivity of the m^{th} measurement at the spatial coordinate \mathbf{r} , and $f(\mathbf{r})$ is the continuous description of the object. The overline in $\bar{\mathbf{g}}$ indicates that the data is averaged over the measurement noise in the imaging system. An alternative short-hand form represents the system as an operator that maps a function of a continuous variable to a finite-dimensional vector; now the imaging equation can be compactly written as

$$\bar{\mathbf{g}}(\mathbf{f}) = \mathcal{H}\mathbf{f}. \quad (12)$$

Here, \mathcal{H} is a continuous-to-discrete linear operator, and \mathbf{f} is an infinite-dimensional vector representing the object [1]. Because it is notationally compact, this form will be used for the remainder of the paper.

3.2. The object model

The quantities we wish to estimate are the elements of a vector called θ . This vector quantifies features or attributes of the object that we would like to estimate, such as signal location, size, and amplitude. Consider an object equation given by

$$f(\mathbf{r}; \theta) = f^{\text{sig}}(\mathbf{r}; \theta) + f^{\text{bknd}}(\mathbf{r}). \quad (13)$$

Here, an object $f(\mathbf{r}; \theta)$ is constructed by adding together a signal $f^{\text{sig}}(\mathbf{r}; \theta)$ and a background $f^{\text{bknd}}(\mathbf{r})$ term. The object is a doubly stochastic random function because both the statistics

on the signal and the statistics on the background determine the overall object statistics. The signal is a deterministic function of the random vector θ . The background is not a function of θ because it contains all of the object structure that is not relevant to the features we seek to estimate. We will proceed with a model of statistically independent signals and backgrounds. The data will be treated as a triply stochastic variable: object fluctuations due to both signal and background variability and measurement noise.

3.3. Measurement noise

When an object is measured by an imaging system, there are fluctuations in the measured values. In a real imaging system, repeatedly acquiring data from the same object will not yield the same image each time. The system sensitivity functions are the deterministic portion of the imaging system, and the noise can be modeled by the addition of a stochastic term whose mean is zero,

$$\mathbf{g} = \mathcal{H}\mathbf{f} + \mathbf{n}. \quad (14)$$

The vector \mathbf{n} is a random perturbation to each data element, whose statistics may in general be object dependent. Many imaging systems operate as photon-counting devices, which leads to a Poisson probability law on the measurement noise of the image data; consequently, for a given object, the noise values in different detector elements are independent of one another. Therefore the noise covariance matrix is diagonal

$$K_{\mathbf{n}} = \langle (\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})) (\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f}))^t \rangle_{\mathbf{g}|\mathbf{f}} = \text{Diag}(\bar{\mathbf{g}}(\mathbf{f})). \quad (15)$$

Here $\text{Diag}(\cdot)$ denotes a diagonal matrix with the vector argument as the diagonal elements, and $\bar{\mathbf{g}}(\mathbf{f})$ is the noise-free image of a given signal plus a given background.

We will later make use of the fact that the noise covariance, when averaged over an object ensemble, remains diagonal

$$\bar{K}_{\mathbf{n}} \equiv \langle K_{\mathbf{n}} \rangle_{\mathbf{f}} = \text{Diag}(\bar{\bar{\mathbf{g}}}). \quad (16)$$

Here $\bar{\bar{\mathbf{g}}}$ denotes the grand mean of the data, defined as the expected value with respect to all sources of randomness in the problem: noise, signal, and background. The object variabilities due to signal and background variations are both accounted for by the average over \mathbf{f} .

4. Image statistics

4.1. Conditional data averages

For the remainder of this paper, the vector \mathbf{g} will denote a noisy image of a random background plus a random signal. A noise-free image of this random object is denoted by

$$\bar{\mathbf{g}}(\mathbf{f}) \equiv \langle \mathbf{g} \rangle_{\mathbf{g}|\mathbf{f}} = \mathcal{H}f^{\text{sig}}(\mathbf{r}; \theta) + \mathcal{H}f^{\text{bknd}}(\mathbf{r}), \quad (17)$$

and the grand mean is a triply stochastic average over detector noise, backgrounds, and signals

$$\bar{\bar{\mathbf{g}}} \equiv \langle \bar{\mathbf{g}}(\mathbf{f}) \rangle_{\mathbf{f}} \equiv \langle \bar{\mathbf{g}}(\mathbf{f}) \rangle_{\theta, \mathbf{f}, \text{bknd}} = \mathcal{H}\bar{f}^{\text{sig}}(\mathbf{r}) + \mathcal{H}\bar{f}^{\text{bknd}}(\mathbf{r}). \quad (18)$$

Rather than adding a third overbar to the grand mean, we define a conditional doubly stochastic average of a noise-free image over just the random background as

$$\bar{\bar{\mathbf{g}}}(\theta) \equiv \langle \bar{\mathbf{g}}(\mathbf{f}) \rangle_{\mathbf{f}, \text{bknd}} = \mathcal{H}f^{\text{sig}}(\mathbf{r}; \theta) + \mathcal{H}\bar{f}^{\text{bknd}}(\mathbf{r}). \quad (19)$$

4.2. Data covariance

A primary ingredient for the optimal linear estimator is the overall data covariance matrix $K_{\mathbf{g}}$, defined by

$$K_{\mathbf{g}} \equiv \left\langle \left\langle (\mathbf{g} - \bar{\mathbf{g}}) (\mathbf{g} - \bar{\mathbf{g}})^t \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\theta, \mathbf{f}^{\text{bknd}}} \quad (20)$$

where $\bar{\mathbf{g}}$ is the grand mean defined in Eq. (18), the average over \mathbf{g} given \mathbf{f} accounts for measurement noise, θ for signal randomness, and \mathbf{f}^{bknd} for background randomness. The data are triply stochastic, and we seek a decomposition of the overall covariance matrix that represents each contributing factor [6]. This can be accomplished by some algebraic manipulation to create uncorrelated random variables that will yield no cross-terms. Consider adding and subtracting the noise-free image of a given object

$$\begin{aligned} K_{\mathbf{g}} &= \left\langle \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})] + [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}] \right\rangle_{\mathbf{g}|\mathbf{f}} \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})] + [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}] \right\rangle_{\mathbf{g}|\mathbf{f}}^t \right\rangle_{\mathbf{f}} \\ &= \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})] [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})]^t \right\rangle_{\mathbf{g}|\mathbf{f}} + \left\langle [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}] [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}]^t \right\rangle_{\mathbf{g}|\mathbf{f}} \\ &= \bar{K}_{\mathbf{n}} + \left\langle [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}] [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}]^t \right\rangle_{\mathbf{g}|\mathbf{f}}, \end{aligned} \quad (21)$$

where the first term has been recognized as the average noise covariance (see Eq. (16)). There are no cross-terms in the above expression because of the uncorrelated nature of the constructed random variables, *i.e.*,

$$\begin{aligned} \left\langle \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})] [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}]^t \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\mathbf{f}} &= \left\langle [\langle \mathbf{g} \rangle_{\mathbf{g}|\mathbf{f}} - \bar{\mathbf{g}}(\mathbf{f})] [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}]^t \right\rangle_{\mathbf{f}} \\ &= \left\langle [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}(\mathbf{f})] [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}]^t \right\rangle_{\mathbf{f}} \\ &= 0. \end{aligned} \quad (22)$$

The first equality is a consequence of all terms except \mathbf{g} being noise-free. Notice that this treatment has not required assumptions about the independence of the individual random processes; instead, it has been designed to create uncorrelated random variables by adding and subtracting conditional averages. This trick can be repeated by adding and subtracting $\bar{\mathbf{g}}(\theta)$ (the image averaged over noise and backgrounds) from the second term in Eq. (21) to decouple the background and signal dependence on the overall covariance. After these manipulations, the overall covariance can be written as

$$K_{\mathbf{g}} = \bar{K}_{\mathbf{n}} + \bar{K}_{\bar{\mathbf{g}}}^{\text{bknd}} + K_{\bar{\mathbf{g}}}^{\text{sig}}. \quad (23)$$

Here, the contribution from the background is

$$\begin{aligned} \bar{K}_{\bar{\mathbf{g}}}^{\text{bknd}} &= \left\langle [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}(\theta)] [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\mathbf{g}}(\theta)]^t \right\rangle_{\mathbf{f}^{\text{bknd}}, \theta} \\ &= \mathcal{H} \left\langle [\mathbf{f}^{\text{bknd}} - \bar{\mathbf{f}}^{\text{bknd}}] [\mathbf{f}^{\text{bknd}} - \bar{\mathbf{f}}^{\text{bknd}}]^{\dagger} \right\rangle_{\mathbf{f}^{\text{bknd}}} \mathcal{H}^{\dagger} \\ &= \mathcal{H} K_{\mathbf{f}}^{\text{bknd}} \mathcal{H}^{\dagger} \end{aligned} \quad (24)$$

where $K_{\mathbf{f}}^{\text{bknd}}$ is the covariance operator for the continuous random variable \mathbf{f}^{bknd} , and \mathcal{H}^{\dagger} is the adjoint of the continuous-to-discrete imaging operator introduced in Section 3.1. Similarly, the signal portion is

$$K_{\bar{\mathbf{g}}}^{\text{sig}} = \left\langle [\bar{\mathbf{g}}(\theta) - \bar{\mathbf{g}}] [\bar{\mathbf{g}}(\theta) - \bar{\mathbf{g}}]^t \right\rangle_{\mathbf{f}^{\text{bknd}}, \theta}$$

$$\begin{aligned}
&= \mathcal{H} \left\langle \left[\mathbf{f}^{\text{sig}} - \bar{\mathbf{f}}^{\text{sig}} \right] \left[\mathbf{f}^{\text{sig}} - \bar{\mathbf{f}}^{\text{sig}} \right]^\dagger \right\rangle_\theta \mathcal{H}^\dagger \\
&= \mathcal{H} \mathbf{K}_f^{\text{sig}} \mathcal{H}^\dagger.
\end{aligned} \tag{25}$$

For Poisson statistics, the noise covariance is diagonal and is given by

$$\bar{\mathbf{K}}_{\mathbf{n}} = \langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})] [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})]^t \rangle_{\mathbf{g}, \mathbf{f}} = \text{Diag}(\bar{\mathbf{g}}). \tag{26}$$

The covariance operators $\mathbf{K}_f^{\text{sig}}$ and $\mathbf{K}_f^{\text{bknd}}$ for the object model used in our study will be specified in Section 5.

4.3. Cross-covariance

The cross-covariance is needed to construct the linear Wiener estimation template. The cross-covariance, which expresses sensitivity of the data to changes in the parameters, is calculated by

$$K_{\theta, \mathbf{g}} = \left\langle (\theta - \bar{\theta}) (\mathbf{g} - \bar{\mathbf{g}})^t \right\rangle_{\theta, \mathbf{g}}, \tag{27}$$

where $\bar{\mathbf{g}}$ is the grand-mean image. To simplify the cross-covariance expression, we begin by making use of the fact that the joint expectation can be performed by a conditional average of the data given a particular θ followed by an expectation with respect to the ensemble of θ . That is,

$$\begin{aligned}
K_{\theta, \mathbf{g}} &= \left\langle \left\langle (\theta - \bar{\theta}) (\mathbf{g} - \bar{\mathbf{g}})^t \right\rangle_{\mathbf{g}|\theta} \right\rangle_\theta \\
&= \left\langle (\theta - \bar{\theta}) \left\langle (\mathbf{g} - \bar{\mathbf{g}})^t \right\rangle_{\mathbf{g}|\theta} \right\rangle_\theta \\
&= \left\langle (\theta - \bar{\theta}) (\bar{\mathbf{g}}(\theta) - \bar{\mathbf{g}})^t \right\rangle_\theta
\end{aligned} \tag{28}$$

where $\bar{\mathbf{g}}(\theta)$ is the image average over noise and backgrounds, given in Eq. (19). Further simplification can be made by recognizing that the grand mean is independent of θ ; thus the term $\langle (\theta - \bar{\theta}) \bar{\mathbf{g}} \rangle_\theta = \mathbf{0}$ and the cross-covariance is

$$K_{\theta, \mathbf{g}} = \left\langle (\theta - \bar{\theta}) (\bar{\mathbf{g}}(\theta))^t \right\rangle_\theta. \tag{29}$$

The system operator can be factored out of the expectation

$$K_{\theta, \mathbf{g}} = \left\langle (\theta - \bar{\theta}) \left(\mathbf{f}^{\text{sig}}(\theta) + \bar{\mathbf{f}}^{\text{bknd}} \right)^t \right\rangle_\theta \mathcal{H}^\dagger. \tag{30}$$

Here, $\mathbf{f}^{\text{sig}}(\theta)$ is a signal parameterized by θ , $\bar{\mathbf{f}}^{\text{bknd}}$ is the average background, and \mathcal{H}^\dagger is the adjoint of the continuous-to-discrete imaging operator. This expression for the cross-covariance is advantageous for studies of system comparison; if the expectation can be evaluated prior to applying the imaging operator, then computational redundancy can be avoided. The background term is a nuisance parameter; it is independent of the parameters θ and therefore will not contribute to the cross-covariance, so

$$K_{\theta, \mathbf{g}} = \left\langle (\theta - \bar{\theta}) \mathbf{f}^{\text{sig}}(\theta)^t \right\rangle_\theta \mathcal{H}^\dagger. \tag{31}$$

In the Wiener estimator expression of Eq. (4), the random background affects only the grand-mean image and the data covariance. The estimates will be biased for a particular object background, but this bias is designed to average to zero by subtracting the grand mean from the data. This is a primary feature of the way in which the optimal linear estimator, the Wiener estimator, treats nuisance parameters.

5. Simulated objects

Thus far, we have provided formulas that are entirely general with respect to the object's signal and background description. To proceed with the calculation of the estimation rules, we must specify a functional form of the continuous object $f(\mathbf{r})$ and the distribution of the object ensemble.

5.1. Parameters of interest: the signal model

The signal portion of the object describes the features we are interested in estimating from the image data. For the purposes of simulations to follow, we confine our treatment to signals that are related to θ by the parameterized signal model, specified in two dimensions as

$$f^{\text{sig}}(\mathbf{r}; \theta) = \frac{A}{I_{\text{cyl}}(R)} \text{cyl}\left(\frac{\mathbf{r} - \mathbf{c}}{R}\right) \quad (32)$$

where

$$\text{cyl}\left(\frac{\mathbf{r} - \mathbf{c}}{R}\right) = \begin{cases} 0, & \text{for } |\mathbf{r} - \mathbf{c}| > R \\ 1, & \text{otherwise} \end{cases} \quad (33)$$

Here, the signal's shape is defined by a circle of radius R , A is a scalar quantity representing the integrated output of the circular signal, $I_{\text{cyl}}(R)$ is the area of the circle $\int d\mathbf{r} \text{cyl}\left(\frac{\mathbf{r} - \mathbf{c}}{R}\right)$, and \mathbf{c} is the location of the signal's center. So a two-dimensional signal model is parameterized by θ that is 5×1 and given by $\theta = (A, R, c_x, c_y)^t$. The signal's spatial distribution is uniquely specified by the vector θ .

In order to reproduce signal variability, the parameter vector θ is treated as a random variable and, consequently, our parameterized object model describes an ensemble of possible $f^{\text{sig}}(\mathbf{r}; \theta)$ distributions. The first- and second- order statistics of this stochastic process are defined as

$$\bar{f}^{\text{sig}}(\mathbf{r}) = \left\langle f^{\text{sig}}(\mathbf{r}; \theta) \right\rangle_{\theta} \quad (34)$$

and

$$K_f^{\text{sig}}(\mathbf{r}, \mathbf{r}') = \left\langle \left(f^{\text{sig}}(\mathbf{r}; \theta) - \bar{f}^{\text{sig}}(\mathbf{r}; \theta) \right) \left(f^{\text{sig}}(\mathbf{r}'; \theta) - \bar{f}^{\text{sig}}(\mathbf{r}'; \theta) \right) \right\rangle_{\theta}. \quad (35)$$

Calculating these expectations requires prior knowledge of the distribution on the parameters $\text{pr}(\theta)$. We will assume that the parameters to be estimated are statistically independent, so the multivariate joint probability density distribution is separable leading to $\text{pr}_{\theta}(\theta) = \text{pr}_A(A) \text{pr}_R(R) \text{pr}_{\mathbf{c}}(\mathbf{c})$.

The signal amplitude is a value naturally constrained to be positive; thus, a shifted gamma distribution is chosen to ensure positivity. The shape of the gamma distribution is determined by the amount of shift and two independent parameters, denoted α_A and β_A

$$\text{pr}_A(A) = \begin{cases} \frac{(A - A_0)^{\alpha_A - 1} \exp\left(\frac{-(A - A_0)}{\beta_A}\right)}{\beta_A^{\alpha_A} \Gamma(\alpha_A)}; & \text{for } A > A_0 \\ 0; & \text{otherwise.} \end{cases} \quad (36)$$

The mean and variance are functions of these parameters and are given by $\bar{A} = \alpha_A \beta_A + A_0$ and $\sigma_A^2 = \alpha_A \beta_A^2$. The mean value of the signal amplitude was determined by constraining the

contrast with respect to the background to be within a realistic range. A normalized histogram for 10,000 random samples drawn from this gamma distribution, $\alpha_A = 4$, $\beta_A = \frac{1}{2}$, and $A_0 = 4.5$ is superimposed on the theoretical pdf in Fig. 1(a). A shifted-gamma distribution was also used

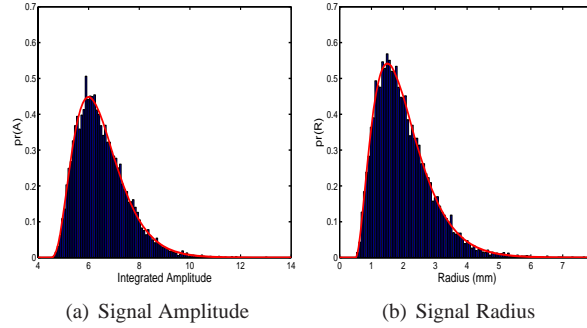


Fig. 1. The probability density function (red line) is superimposed on a histogram of 10,000 samples.

to describe the variability of the signal radius R

$$\text{pr}_R(R) = \begin{cases} \frac{(R-R_0)^{\alpha_R-1} \exp\left(\frac{-(R-R_0)}{\beta_R}\right)}{\beta_R^{\alpha_R} \Gamma(\alpha_R)}; & \text{for } R > R_0 \\ 0; & \text{otherwise.} \end{cases} \quad (37)$$

The gamma distribution has been shifted such that all signal radii have a lower bound equal to the spacing of measured voxel points; $R_0 = 0.5$ mm. Without this constraint, all radii below 0.5 mm would produce the same object and the parameter would be unidentifiable from even noise-free images. For all studies that follow, $\bar{R} = 2.0$ mm, $\alpha = 3$, and $\beta = 0.5$ mm. Figure 1(b) is a plot of a gamma pdf with these specified parameters along with the frequency of occurrence for 10,000 samples drawn from this distribution.

The prior distribution on the location of the signal is a multivariate uncorrelated Gaussian centered at the origin. In our simulation study, the linear dimension of the object space field of view is 30 mm, and the width parameter of the Gaussian distribution is set to $\sigma_c = 4$ mm, which limits the probability of signal-absent images.

5.2. Nuisance parameters: the background model

A lumpy background model, developed by Rolland and Barrett [18], consists of a summation of lumps superimposed within the field of view (FOV). The functional form of the model is

$$f^{\text{bkgnd}}(\mathbf{r}) = S(\mathbf{r}) \left(A_b + \sum_{n=1}^L l(\mathbf{r} - \mathbf{r}_n) \right). \quad (38)$$

The function $S(\mathbf{r})$ represents the support region, which is unity within the imaging system's FOV and zero otherwise, A_b is the uniform contribution to the background and $l(\mathbf{r})$ is the functional description of each lump.

The total number of lumps and the location of each lump are the random quantities that determine the stochastic nature of this background model. The number of lumps within each realization of the background, denoted L in Eq. (38), is a Poisson-distributed random variable.

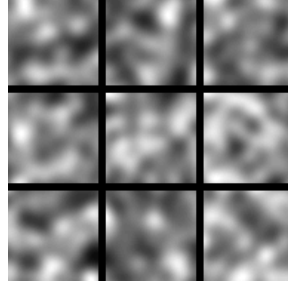


Fig. 2. Samples of 2D Gaussian lumpy backgrounds.

The centroid of each lump, r_n , is uniformly distributed throughout the FOV. This uniformity leads to a mean lumpy background that is constant within the support region and given by

$$\bar{f}^{\text{bkgnd}}(\mathbf{r}) = \left\langle f^{\text{bkgnd}}(\mathbf{r}) \right\rangle_{\mathbf{r}_{n,L}} = S(\mathbf{r}) \left(A_b + \frac{\bar{L} I_l}{V_{FOV}} \right). \quad (39)$$

Here, I_l is the area of a single lump $\int d\mathbf{r} l(\mathbf{r})$, \bar{L} is the average number of lumps, and V_{FOV} is the volume of the FOV.

The autocovariance is a second-order statistic that requires evaluating the expectation of the random process at two distinct points. In general, the autocovariance of a stochastic process is a function of this pair of points

$$K_f^{\text{bkgnd}}(\mathbf{r}, \mathbf{r}') = \left\langle f^{\text{bkgnd}}(\mathbf{r}) f^{\text{bkgnd}}(\mathbf{r}') \right\rangle_{\mathbf{r}_{n,L}} - \bar{f}^{\text{bkgnd}}(\mathbf{r}) \bar{f}^{\text{bkgnd}}(\mathbf{r}'). \quad (40)$$

The lumpy background model is a special case of a filtered Poisson point process [1]. In the theoretical limit of infinite support, equivalent to the case of $S(\mathbf{r}) = 1$ everywhere, the first- and second-order statistics of the lumpy background model satisfy stationarity conditions. In practice, stationarity is inhibited by the effect of finite support. The autocovariance for a lumpy background within a finite FOV retains the hallmark of stationarity, being functionally dependent on the difference between a pair of points, but the support region truncates the otherwise stationary function:

$$K_f^{\text{bkgnd}}(\mathbf{r}, \mathbf{r}') = S(\mathbf{r}) S(\mathbf{r}') \left(\frac{\bar{L}}{V_{FOV}} [l \otimes l](\mathbf{r} - \mathbf{r}') \right), \quad (41)$$

where $[l \otimes l](\mathbf{r} - \mathbf{r}')$ represents the autocorrelation of the lump function evaluated at $\mathbf{r} - \mathbf{r}'$. In this study, the lump function $l(\mathbf{r})$ is chosen to be Gaussian $l(\mathbf{r}) = b_0 \exp(-\mathbf{r}^2/2\sigma^2)$. The width parameter σ and peak value b_0 of each Gaussian are selected to produce lumps that are broad and low amplitude with respect to the signal. Nine samples from this background ensemble are evaluated on a discrete planar grid and displayed in Fig. 2.

The first- and second-order statistics of this Gaussian lumpy background model are

$$\bar{f}^{\text{bkgnd}}(\mathbf{r}) = S(\mathbf{r}) \left(A_b + \frac{\bar{L} b_0 (2\pi)^{3/2} \sigma^3}{V_{FOV}} \right) \quad (42)$$

and

$$K_f^{\text{bkgnd}}(\mathbf{r}, \mathbf{r}') = S(\mathbf{r}) S(\mathbf{r}') \left(\frac{\bar{L} b_0^2 \pi^{3/2} \sigma^3 \exp\left(\frac{-(\mathbf{r}-\mathbf{r}')^2}{4\sigma^2}\right)}{V_{FOV}} \right). \quad (43)$$

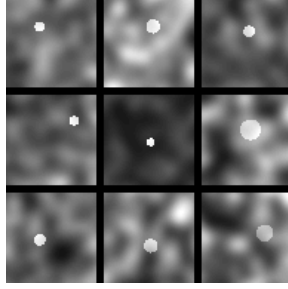


Fig. 3. Noise-free images generated by sampling from the ensemble of possible signals and backgrounds.

These first- and second-order statistics have been calculated using the distribution on lump location and number of lumps; note that the pdf $\text{pr}(\mathbf{f}^{\text{bkgnd}})$ has not been specified and is not necessarily Gaussian.

6. Simulated images and image statistics

Typically, a simulation study requires a careful trade-off between the realism of the model and the computational complexity of the implementation. The simulation of real imaging systems is usually informed by knowledge of the acquisition geometry and/or measurement data from phantoms designed to probe the system's performance. Our goal however, is to evaluate the performance of the proposed estimators in very simplistic test cases. Our simulated imaging system is overly optimistic in modeling the transfer of information from object to image. We will demonstrate some cases for which linear estimation fails in this simulation, thereby providing evidence that it would not work in a realistic imaging context. The scanning-linear estimator will also be evaluated for this simple case, and the successful results motivate future work beyond the current simulation.

Consider a sampling operator given by

$$[\bar{\mathbf{g}}]_m = f(\mathbf{r}_m), \quad (44)$$

where $f(\mathbf{r})$ is a planar object, each \mathbf{r}_m is a spatial coordinate on an evenly spaced grid, and m is a lexicographic ordering index. In our study, the planar object is sampled on a 61×61 grid of 0.5 mm spacing resulting in a simulated image with a linear FOV of 30 mm in each direction.

To convey a sense of the amount of signal and background variability present in the simulated data, nine realizations of noise-free data are displayed in Fig. 3. These images are generated by drawing samples from the signal and background ensembles, described in Section 5.2 and Section 5.1 respectively.

The mean image of the signal and background is 131,000 detected events, *i.e.*, $\sum_m \bar{\mathbf{g}}_m \approx 131,000$. The simulated detector face contains 61^2 pixels; therefore the average count density is approximately 35 events per pixel. Poisson noise is added to each image and the average density indicates the relative contribution of the measurement noise (*i.e.*, photon-counting noise) to the overall stochastic fluctuation of the data. Figure 4 shows nine noisy images.

6.1. Calculating the cross-covariance

The expectation of the product of the parameterized signal model and the parameter vector, introduced in Eq. (31), is most naturally calculated one row at a time, where the decomposition

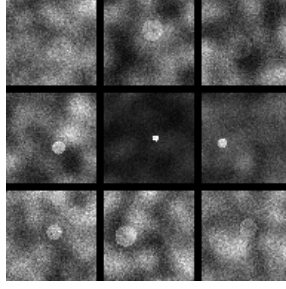


Fig. 4. Noisy images generated by sampling from the ensemble of possible signals and backgrounds, and adding Poisson noise.

is given by

$$\left\langle (\theta - \bar{\theta}) \mathbf{f}^{\text{sig}}(\theta)^t \right\rangle_{\theta} = \begin{bmatrix} \left\langle (A - \bar{A}) \mathbf{f}^{\text{sig}}(\theta)^t \right\rangle_{\theta} \\ \left\langle (R - \bar{R}) \mathbf{f}^{\text{sig}}(\theta)^t \right\rangle_{\theta} \\ \left\langle (c_x - \bar{c}_x) \mathbf{f}^{\text{sig}}(\theta)^t \right\rangle_{\theta} \\ \left\langle (c_y - \bar{c}_y) \mathbf{f}^{\text{sig}}(\theta)^t \right\rangle_{\theta} \end{bmatrix}. \quad (45)$$

Here, the definition of the vector $\theta \equiv [A, R, c_x, c_y]$ has been used, and each row is an infinite-dimensional vector in object space. Consider the functional form of the row corresponding to signal amplitude

$$\begin{aligned} \left\langle (A - \bar{A}) f^{\text{sig}}(\mathbf{r}; \theta) \right\rangle_{\theta} &= \int d\mathbf{c} \text{pr}(\mathbf{c}) \int dR \text{pr}(R) \times \\ &\quad \int dA \text{pr}(A) (A - \bar{A}) \frac{A}{I_{\text{cyl}}(R)} \text{cyl} \left(\frac{\mathbf{r} - \mathbf{c}}{R} \right) \\ &= \frac{\sigma_A^2}{\bar{A}} \bar{f}^{\text{sig}}(\mathbf{r}). \end{aligned} \quad (46)$$

Here $\bar{f}^{\text{sig}}(\mathbf{r})$ is the average signal, given in Eq. (34), $I_{\text{cyl}}(R)$ is the area of the signal function, and the variance of A is given by $\sigma_A^2 = \int dA \text{pr}(A) A (A - \bar{A})$. The amplitude is the only parameter linearly related to the object, so the expectation is simply proportional to the average signal. The integral expressions for the other rows are closely related to one another and can be written as

$$\left\langle (R - \bar{R}) f^{\text{sig}}(\mathbf{r}; \theta) \right\rangle_{\theta} = \bar{A} \left\langle \int d\mathbf{c} \text{Gaus}(\mathbf{c}) \frac{(R - \bar{R})}{I_{\text{cyl}}(R)} \text{cyl} \left(\frac{\mathbf{r} - \mathbf{c}}{R} \right) \right\rangle_R, \quad (47)$$

and

$$\left\langle (c_i - \bar{c}_i) f^{\text{sig}}(\mathbf{r}; \theta)^t \right\rangle_{\theta} = \bar{A} \left\langle \int d\mathbf{c} \text{Gaus}(\mathbf{c}) \frac{(c_i - \bar{c}_i)}{I_{\text{cyl}}(R)} \text{cyl} \left(\frac{\mathbf{r} - \mathbf{c}}{R} \right) \right\rangle_R. \quad (48)$$

Here $\text{Gaus}(\cdot)$ is the two-dimensional Gaussian distribution representing $\text{pr}(\mathbf{c})$. The integration over \mathbf{c} is performed using Gaussian quadrature over the finite support of the signal function. The outer expectation over the radius is evaluated using Monte Carlo integration. Samples are drawn from the univariate distribution $\text{pr}(R)$, and the sample mean of the integral expression is used as an unbiased estimator of this quantity. We used ten million R samples, adding one million samples until the maximum change in the solution was below a thousandth of a percent. The coordinates were chosen using the sampling operator introduced in Eq. (44), so the continuous spatial function, $f^{\text{sig}}(\mathbf{r}; \theta)$, was sampled on an evenly spaced grid of 61×61 points to obtain a discrete image of the signal.

The results of this cross-covariance calculation are displayed in Fig. 5, where each row is displayed as a planar image. Evaluating $K_{\theta, \mathbf{g}}$ at a given row selects one element of the vector θ and measures the cross-covariance between each image pixel and the selected parameter. Viewing the data in this fashion allows for visual interpretation of how the Wiener estimator uses the image data. The cross-covariance for signal amplitude A is provided in Fig. 5(b) and is non-negative for all pixel values because an increase in the parameter A would either increase or not change each pixel value of the image. This template is circularly symmetric about the center of the image, which is the average signal location. The cross-covariance for signal radii is displayed in Fig. 5(a). Adjacent negative and positive lobes are features common to edge-detection algorithms, and the Wiener radius template also has this type of distribution. The radius template is circularly symmetric and centered about the average signal location. The templates for location estimation, 5(c) and 5(d), look like a centroid, or center of mass, calculation that tapers off at the edge of the image where signal is less likely to be located.

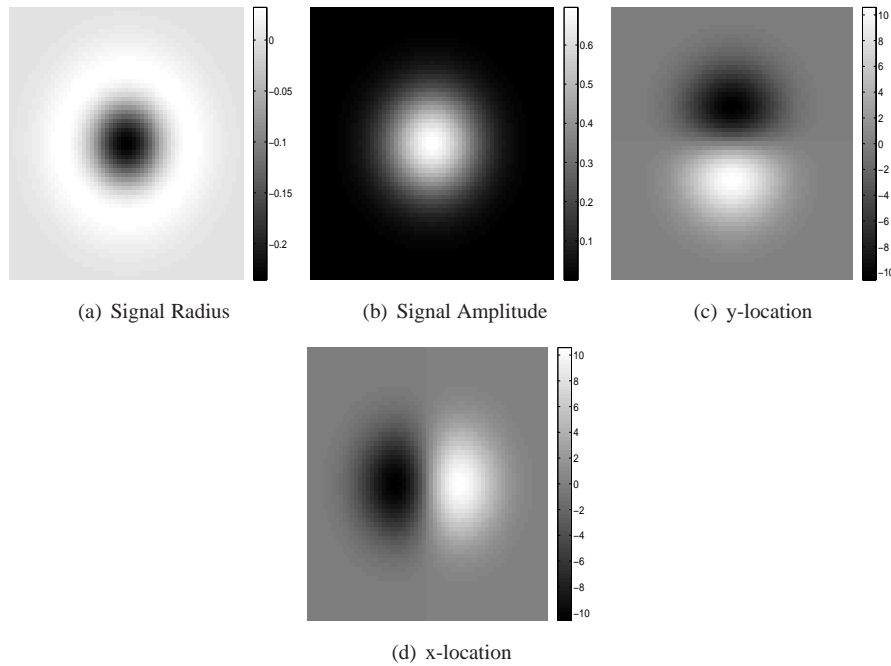


Fig. 5. Rows of the matrix $K_{\theta, \mathbf{g}}$, the cross-covariance between the data and the parameters to be estimated.

6.2. Calculating the data covariance

A data covariance matrix can be difficult to evaluate due to finite data storage. If M measurements are made by the imaging system, the covariance matrix is $M \times M$. For tomographic systems, temporal systems, or any system that collects large amounts of data, explicit storage of each covariance element may not be practical. Potentially useful solutions include a decomposition of a sample covariance matrix [19] or projection onto a lower-dimensional orthonormal set (*i.e.*, channelizing the data). The imaging operator \mathcal{H} of Eq. (44) samples the continuous function on a grid of 61×61 points; consequently the overall data covariance matrix $K_{\mathbf{g}}$ is only $61^2 \times 61^2$, and storage requirements are minimal.

The noise term is known for our model: it is diagonal, and all nonzero elements are given by the noise-free image of the overall average object. Eq. (43) is an analytic expression for the Gaussian lumpy background term, which is transformed through the system operator to data space. Each element of the object parameter term in Eq. (35) can be reduced to an integral expression at each coordinate in object space

$$\left[K_f^{\text{sig}}\right]_{m,m'} = \langle A^2 \rangle_A \left\langle \frac{1}{I_{\text{cyl}}(R)^2} \int d^2c \text{Gaus}(\mathbf{c}) \text{cyl}\left(\frac{\mathbf{r}_m - \mathbf{c}}{R}\right) \text{cyl}\left(\frac{\mathbf{r}_{m'} - \mathbf{c}}{R}\right) \right\rangle_R, \quad (49)$$

where \mathbf{r}_m is a point in object space. This integral expression is evaluated using the same Monte-Carlo and Gaussian quadrature hybrid method employed for the cross-covariance calculation. We used 100,000 samples from the radius distribution. The integral with respect to the signal center \mathbf{c} was evaluated by quadrature at these 100,000 samples, and the sample mean is used as an unbiased estimator of the integral's true value. The sample size of 100,000 was chosen by adding 10,000 samples until the maximum change in the estimates was less than 1/10 of a percent.

The organization of each covariance element into a matrix requires lexicographically organizing the rows and columns of an image into one index. Then, one row (or equivalently one column since all covariance matrices are symmetric) of K_g selects a given pixel location, and the M values of this row are the covariance between the selected pixel and all other pixels in the image. The covariance values could also be organized into a function of four indices, the row and column designation for the pair of pixels at which the covariance is evaluated. With this type of organization, the covariance would be written as $K_g(m, n, m', n')$, where (m, n) is the row and column designation of one pixel and (m', n') is the designation of the other. In this fashion $K_g(m, n, :, :)$ is a $J \times J$ matrix where $J^2 = M$, and the colon denotes evaluating the function at all indices. This matrix can be viewed as an image in data space that expresses the covariance of the (m, n) pixel with each of the others. Such an image would peak at the point $K_g(m, n, m, n)$ because this is the variance of the (m, n) pixel.

Covariance is a way to measure trend. When a given pixel increases if another pixel also *tends* to increase, then the covariance will be positive for this pair of pixels. A negative trend yields a negative covariance. Completely uncorrelated pixels will have a covariance of zero. These points are illustrated in Fig. 6. Because the noise is independent, the images $\bar{K}_g^{\text{noise}}(m, n, :, :)$, displayed in Figs. 6(a) and 6(d), will be zero except at the point (m, n) . The background covariance is stationary within the FOV, and in the corresponding images of $\bar{K}_g^{\text{bkgn}}(m, n, :, :)$ (Figs. 6(b) and 6(e)), the only feature that changes is the centroid of the correlation function. The results for the signal covariance is shown in Fig. 6(c) and 6(f). This covariance expression evaluates the area of overlap between two signals, weighted by a Gaussian distribution in the center. The contribution from the Gaussian disturbs circular symmetry and results in a higher covariance function near the center of the image.

6.3. Calculating the scanning linear estimator

In Section 2.2, we introduced a scanning-linear estimator derived from MAP estimation with two critical approximations to the data likelihood: a Gaussian distribution and a covariance independent of θ . In this section, we will use the functional form of the object model introduced thus far to derive the task-specific scanning-linear estimator.

The first assumption is that the likelihood of the data conditioned on the parameters, called $\text{pr}(\mathbf{g}|\theta)$, is Gaussian. For a majority of realistic imaging problems, the noise-free data will not be uniquely determined by the parameters to be estimated because of nuisance parameters. In

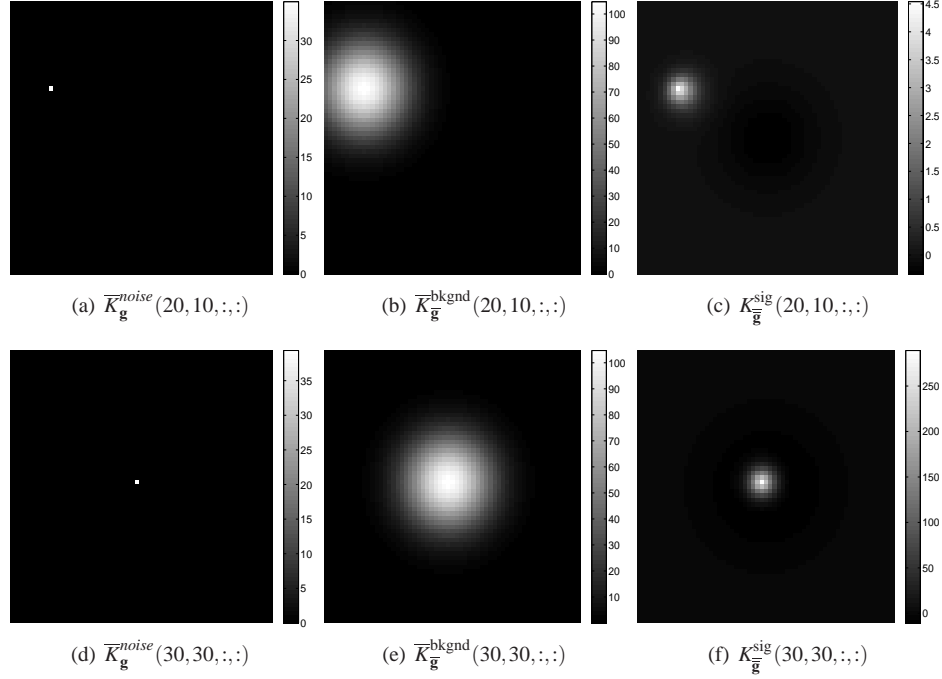


Fig. 6. The three terms that contribute to the triply stochastic data covariance $K_{\mathbf{g}}$.

these circumstances, the likelihood is given by a marginalization over the nuisance parameters

$$\text{pr}(\mathbf{g}|\theta) = \int d^M \mathbf{b} \text{pr}(\mathbf{g}|\theta; \mathbf{b}) \text{pr}(\mathbf{b}) \quad (50)$$

where $\mathbf{b} = \mathcal{H}\mathbf{f}^{\text{bknd}}$, the noise-free image of a background. The measurement noise is described by $\text{pr}(\mathbf{g}|\theta; \mathbf{b})$ the probability of an image when the object is fixed. Maximum-likelihood estimation requires numerous evaluations of the integral expression in Eq. (50) to find the value of θ that maximizes the likelihood. Then this search, and subsequent integral evaluations, need to be repeated for each image \mathbf{g} . Avoiding this calculation is the utility of assuming a Gaussian distributed likelihood. The mean of the approximated likelihood is simply the data averaged over backgrounds and noise but conditioned on θ

$$\bar{\mathbf{g}}(\theta) = \mathcal{H}\mathbf{f}^{\text{sig}}(\theta) + \mathcal{H}\bar{\mathbf{f}}^{\text{bknd}} \quad (51)$$

$$\equiv \mathbf{s}(\theta) + \bar{\mathbf{b}} \quad (52)$$

where $\mathbf{s}(\theta)$ is the noise-free image of a signal parameterized by θ , and $\bar{\mathbf{b}}$ is the noise-free image of the mean background. The conditional covariance only contains terms for the noise and the background

$$\begin{aligned} K_{\mathbf{g}|\theta} &= \left\langle (\mathbf{g}(\theta) - \bar{\mathbf{g}}(\theta)) (\mathbf{g}(\theta) - \bar{\mathbf{g}}(\theta))^t \right\rangle_{\mathbf{g}|\theta} \\ &= \text{Diag}(\bar{\mathbf{g}}(\theta)) + \bar{K}_{\bar{\mathbf{g}}}^{\text{bknd}}. \end{aligned} \quad (53)$$

The second approximation is to replace the conditional covariance with the ensemble averaged covariance to avoid the functional dependence on θ . These matrices only differ in the noise

term, so replacing the conditional covariance by its average is equivalent to making a slowly varying signal approximation in the noise covariance

$$K_{\mathbf{g}|\theta} \approx \langle K_{\mathbf{g}|\theta} \rangle_{\theta} = \text{Diag}(\bar{\mathbf{g}}) + \bar{K}_{\mathbf{g}}^{\text{bknd}}. \quad (54)$$

Substituting these expressions into Eq. (10) shows that the scanning-linear estimator is calculated by optimizing the function

$$\hat{\theta}_{\text{SL}}(\mathbf{g}) = \underset{\theta}{\text{argmax}} \left\{ \mathbf{s}(\theta)^T \left(\text{Diag}(\bar{\mathbf{g}}) + \bar{K}_{\mathbf{g}}^{\text{bknd}} \right)^{-1} \left(\mathbf{g} - \bar{\mathbf{b}} - \frac{1}{2} \mathbf{s}(\theta) \right) + \ln[\text{pr}(\theta)] \right\}. \quad (55)$$

The entire argument within $\text{argmax}\{\cdot\}$ is called the objective function, and the scanning linear estimate is the one that maximizes this objective function. For this problem, the objective function is a continuous scalar-valued function of the parameters to be estimated as well as a linear transform of the image data. Therefore, the optimization is performed with respect to the continuous vector θ for each image \mathbf{g} .

7. Results

7.1. Estimating signal location, radius, and amplitude

The signal location, radius, and amplitude are all random variables, and the task is to estimate these quantities from noisy image data acquired when the signals are embedded in a Gaussian lumpy background.

7.1.1. Wiener results

Ten thousand simulated images were used to generate Wiener estimates. Each element of $\hat{\theta}$ (amplitude, radius, and location) is shown in Fig. 7. As this figure illustrates, the Wiener estimator is consistently returning nearly the same value regardless of the image data. The Wiener estimator fails to utilize the image data and instead merely returns the prior mean on the parameters. Each element of $\hat{\theta}$ hovering around $\bar{\theta}$ is shown in Fig. 7. This behavior is due to the relative magnitude of the two terms that constitute the Wiener estimator (see Eq. (4)): one is a linear operation on the deviation of the image data from the grand mean, and the second adds the prior mean $\bar{\theta}$. If the estimator is averaged over the image ensemble, the first term mentioned is zero, and the average estimate is equal to the average of the parameters, a trait known as globally unbiased estimation. When the signal location, radius, and shape are unknown the Wiener estimator is the worst possible globally unbiased estimator. The linear operations on the data do not just average to zero, but are nearly zero for all images. Instead of using the image to inform the estimation, it is effectively ignored, and an estimate close to the prior mean θ is returned.

7.1.2. Scanning linear results

A thorough search of θ space is not possible due to its continuous nature; instead, possible solutions are discretized to a grid of approximately 14 million points. These points are generated by spacing the location solutions a voxel apart, for a total of 61×61 locations. The radius is incremented by one voxel for a total of 16 candidate radii. Twenty amplitude solutions are evenly spaced over a plausible range. These candidate θ solutions are evaluated for 300 sample images. For each image, the maximum returned by the discrete search occurs at a point neighboring the true parameter values. Examples are shown in Fig. 8, where a cross-section of the objective function is paired with the corresponding image data. Values near the true signal radius and amplitude are held constant while the location is scanned across each voxel, and the scalar value of the objective function at these points populates the displays in Fig. 8.

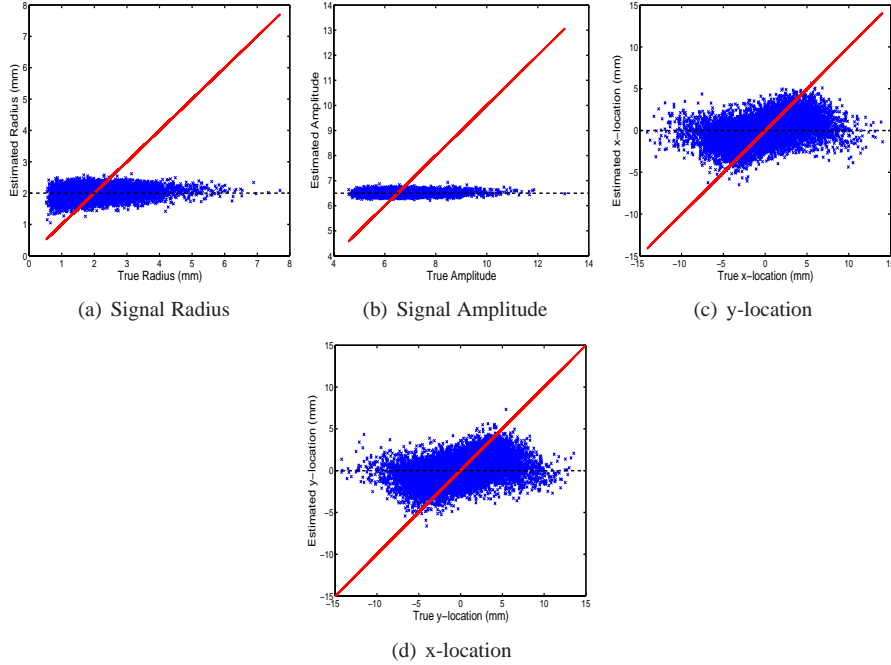


Fig. 7. Scatter plots of the Wiener estimates versus the true parameter values for each element of $\theta = [A, R, c_x, c_y]$. The red line indicates perfect estimation performance, the actual estimates are blue points, and the black line shows $\bar{\theta}$, the prior mean.

Figure 8 shows that the maximum value of the objective function is very near the centroid of the signal in the image. Around this maximum, the objective function has smaller-scale fluctuations that are symptomatic of the pre-whitening operation (K_g^{-1}) in the objective function and the image's noisy pixel values. Conventional optimization methods do not work well when there are local extrema between the algorithms' starting point and the global extremum. To circumvent this obstacle, a simulated annealing routine is started at the maximum found by the grid evaluations. Beginning a simulated annealing search at these proposed solutions further verifies that the final solution is the global maximum of this objective function. The results of the simulated annealing search are plotted, against the true parameter values, for 300 sample images in Fig. 9. The estimate returned by our search method yields a value of the objective function that is either equal to or greater than the function evaluated at the true signal parameters; therefore the search implementation is not responsible for any deviation from perfect estimation performance in Fig. 9. The sample EMSEs are 0.0024 for the radius component, 0.3200 for the amplitude, and 0.0028/0.0033 for each of the location components. The EMSEs are a few orders of magnitude below the prior variance of $\text{pr}(\theta)$ for location and radius. The amplitude estimate is most sensitive to noise and background effects, but the scanning linear EMSE is still about three times lower than an estimate that merely returns the mean.

7.2. Radius of signal known, location, and amplitude to be estimated

In this task, the radius of each signal is fixed at $R = 2$ mm, which corresponds to a diameter of 8 pixels. Ten thousand signals are sampled from the ensemble $\text{pr}(\theta)$, added to samples from the lumpy background distribution, and noisy image data are created. The scanning-linear

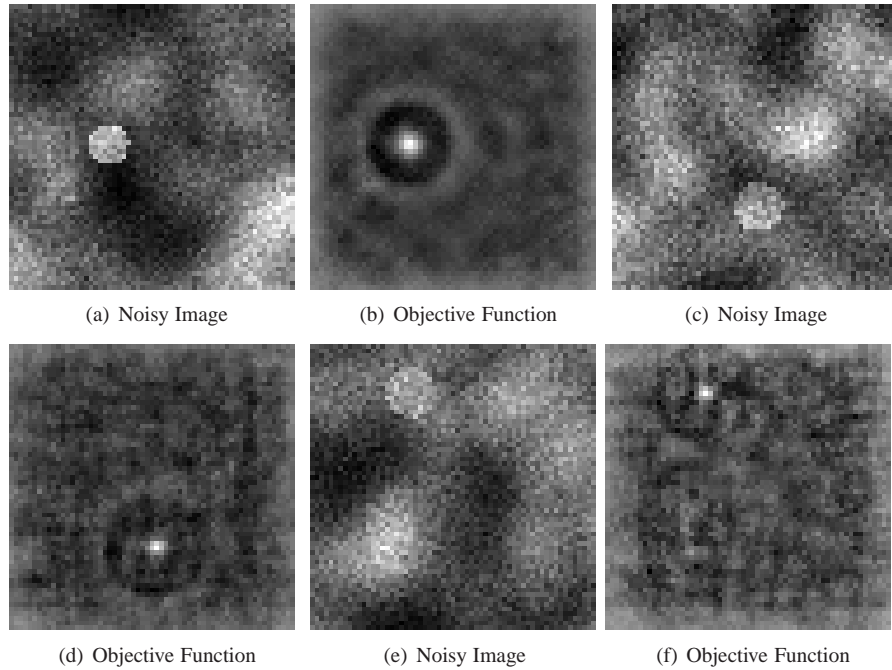


Fig. 8. The objective functions scanned over solutions for signal locations are paired with the noisy image data.

estimator is not implemented for this task; the outcome is anticipated to be redundant because it performed well in the more complex task of radius unknown.

7.2.1. Wiener results

The location and amplitude are estimated from each image \mathbf{g} , and the resulting sample EMSE is calculated for each element of the vector $\hat{\boldsymbol{\theta}}(\mathbf{g}) = [\hat{A}, \hat{\mathbf{c}}_x, \hat{\mathbf{c}}_y]$. We are interested in the effect prior knowledge on location has on the EMSE of the amplitude estimates. The prior distribution on location is a two-dimensional Gaussian; hence shrinking the standard deviation $\sigma_{\mathbf{c}}$ corresponds to less variability in the location of the signals. The sample EMSE of the amplitude estimate as it varies with the standard deviation of signal locations is plotted in Fig. 10(a). It is to be expected that, when the signal location varies less, the amplitude estimates should improve. This is confirmed by the simulation study, but the effect is not substantial.

The smallest $\sigma_{\mathbf{c}}$ included in the study is 1.0 mm (2 pixels), and the estimation performance for this case is plotted in Fig. 11. The location is estimated rather accurately, as shown in Figs. 11(b) and 11(c), but the amplitude estimation given in Fig. 11(a) is not trending well with the data. The sample EMSE of these amplitude estimates is 0.7550 when $\sigma_{\mathbf{c}}$ 1.0 mm (2 pixels), which can be read off of Fig. 10(a). Because the variance of $\text{pr}(A)$ is 1.0, the ratio of Wiener EMSE to intrinsic variance is 75%, indicating only a 25% reduction in EMSE compared to an estimator that simply returns the average amplitude. The performance of the Wiener location estimator has a sample EMSE of 0.10 when $\sigma_{\mathbf{c}}$ is 1.0 mm (2 pixels), so the ratio of sample EMSE to the variance of $\text{pr}(\mathbf{c})$ is 10%. This low ratio indicates the strong performance seen in Figs. 11(b) and 11(c). Instead of looking at scatter plots for a range of location variability the ratio of sample EMSE of the x-location estimate to prior variance $\sigma_{\mathbf{c}}^2$ is plotted in Fig. 10(b). This

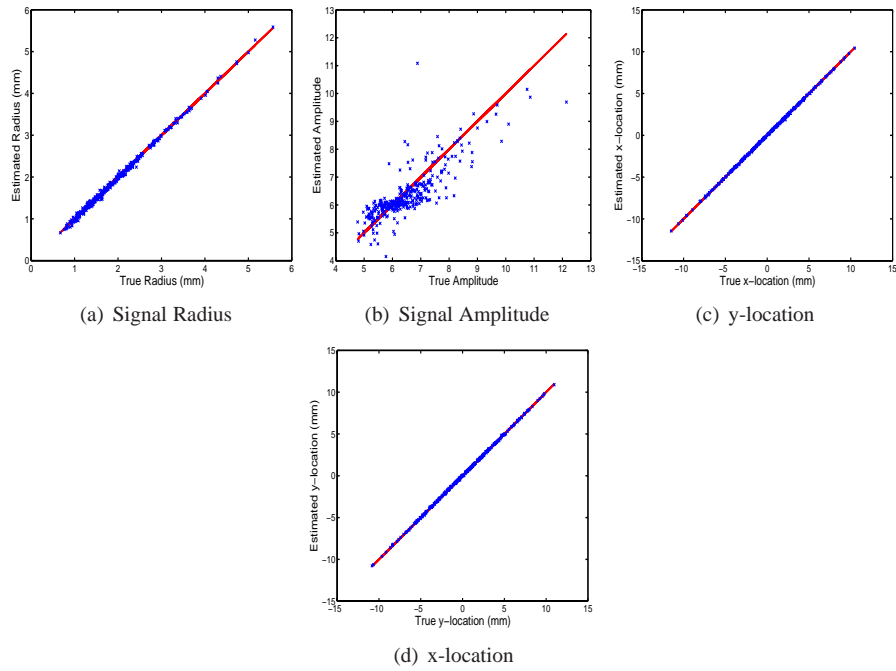


Fig. 9. Scatter plots of the scanning linear estimates versus the true parameter values for each element of $\theta = [A, R, c_x, c_y]$. The red line indicates perfect estimation performance and the actual estimates are blue points.

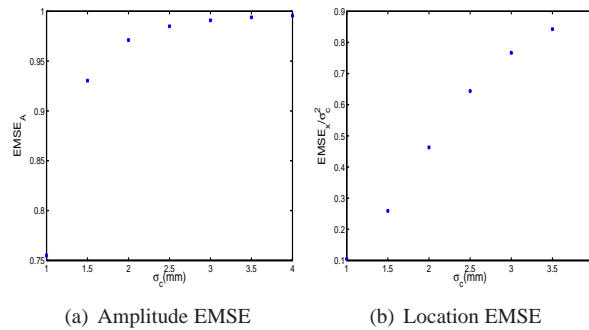


Fig. 10. (a): The Wiener EMSE of the estimated signal amplitude as dependent upon the standard deviation of the pdf on signal location. (b): The Wiener EMSE of the estimated location divided by the variance of the pdf on signal location as dependent upon the width of the pdf on signal location.

measure of performance indicates that, when the signal location is distributed throughout the FOV, linear estimation is not good at finding it; when σ_c is 4.0mm (8 pixels) the EMSE nearly equals the prior variance σ_c^2 , and the performance ratio is close to one. However, when the standard deviation of $\text{pr}(\mathbf{c})$ is 1.50 mm (3 pixels), the Wiener EMSE is reduced by about 75% compared to an estimate that merely guesses the average location. These results suggests that, although Wiener estimation cannot locate signals of known shape when the location variability is large, it has potential when the location is known to be within a specific region.

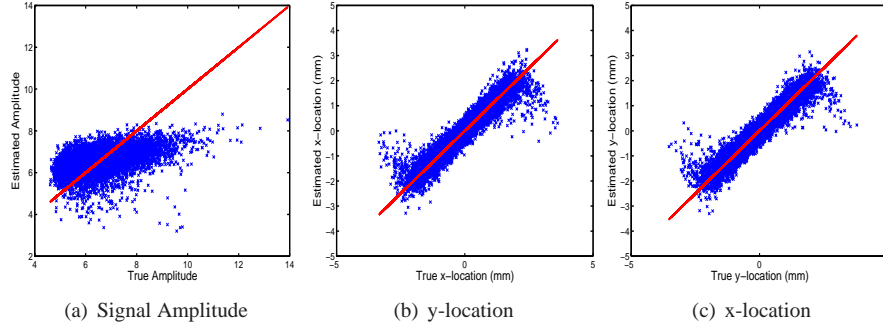


Fig. 11. Scatter plots of the Wiener estimates versus the true parameter values for each element of $\theta = [A, c_x, c_y]$ when the radius is fixed at 2 mm (4 pixels). The red line indicates perfect estimation performance, and the actual estimates are blue points.

7.3. Location and radius of signal known, amplitude to be estimated

This study seeks to estimate the amplitude of a signal when the location and shape are both known, so that the vector of parameters to estimate reduces to a scalar $\theta = A$. The radius is fixed at 2 mm; and the signal is always located at the center of object space. As in previous simulations 10,000 of these signals with varying amplitude are embedded in randomly selected Gaussian lumpy backgrounds, and noisy images are generated.

7.3.1. Wiener results

For the case of location and shape known, the performance of the Wiener estimator is nearly perfect, as shown in the scatter plot of Fig. 12. The sample EMSE is 0.0727, a remarkable improvement over an estimator that simply returns the mean. When the estimate equals the mean, the EMSE equals 1.00, the prior variance of $\text{pr}(A)$.

To provide an understanding of the calculation done by the Wiener estimator, a central slice of the matrix $K_{A,g}K_g^{-1}$ is plotted in Fig. 13, superimposed with a slice of the average signal. These functions have been normalized by their respective maximum values to accentuate their regions of overlap. The estimation template resembles a high-pass filtered version of the signal because this is precisely the action of the matrix multiplication by K_g^{-1} , also called the pre-whitening step.

7.3.2. Scanning linear

For the case of estimating a single scalar parameter that is linearly related to the object, a simple gradient-minimization can be solved to find an approximation to the scanning linear estimator.

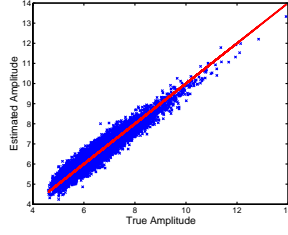


Fig. 12. Scatter plot of true versus Wiener estimates of amplitude when the signal shape and location are known *a priori*. The red line indicates perfect estimation performance, and the actual estimates are blue points.

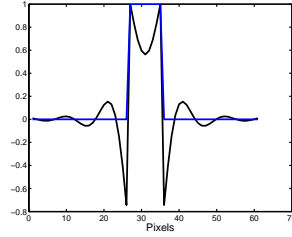


Fig. 13. The linear template used on image data (black line) and the average signal (red line) are normalized by their respective maximum values to accentuate their regions of overlap.

Consider Eq. (55) with the substitution for θ as the scalar amplitude A

$$\hat{A}_{\text{SL}}(\mathbf{g}) = \operatorname{argmax}_A \left\{ \mathbf{s}(A)^t \left(\operatorname{Diag}(\bar{\mathbf{g}}) + \bar{K}_{\mathbf{g}}^{\text{bknd}} \right)^{-1} \left(\mathbf{g} - \bar{\mathbf{b}} - \frac{1}{2} \mathbf{s}(A) \right) + \ln[\operatorname{pr}(A)] \right\} \quad (56)$$

where $\mathbf{s}(A)$ is the noise-free image of a signal with amplitude A . The functional dependence between signal and amplitude is simply linear

$$\mathbf{s}(A) = A\tilde{\mathbf{s}}, \quad (57)$$

where $\tilde{\mathbf{s}}$ is a normalized signal image. This substitution yields

$$\hat{A}_{\text{SL}}(\mathbf{g}) = \operatorname{argmax}_A \left\{ A\tilde{\mathbf{s}}^t \left(\operatorname{Diag}(\bar{\mathbf{g}}) + \bar{K}_{\mathbf{g}}^{\text{bknd}} \right)^{-1} \left(\mathbf{g} - \bar{\mathbf{b}} - \frac{1}{2} A\tilde{\mathbf{s}} \right) + \ln[\operatorname{pr}(A)] \right\}. \quad (58)$$

Without the prior $\operatorname{pr}(A)$, this objective function is quadratic in A , and the maximization could be calculated by a gradient minimization. This modification is akin to maximum-likelihood (ML) estimation; the estimator maximizes the Gaussian approximated likelihood with the slowing-varying covariance assumption of Eq. (54). If the prior is ignored, the remaining terms are maximized at the estimate

$$\hat{A}(\mathbf{g}) = \frac{\tilde{\mathbf{s}}^t \left(\operatorname{Diag}(\bar{\mathbf{g}}) + \bar{K}_{\mathbf{g}}^{\text{bknd}} \right)^{-1} (\mathbf{g} - \bar{\mathbf{b}})}{\tilde{\mathbf{s}}^t \left(\operatorname{Diag}(\bar{\mathbf{g}}) + \bar{K}_{\mathbf{g}}^{\text{bknd}} \right)^{-1} \tilde{\mathbf{s}}} \quad (59)$$

which is unbiased

$$\langle \hat{A}(\mathbf{g}) \rangle_{\mathbf{g}|A} = A \quad (60)$$

because $\langle \mathbf{g} \rangle_{\mathbf{g}|A} = A\tilde{\mathbf{s}} + \bar{\mathbf{b}}$. This estimate is plotted in Fig. 14 for 10,000 sample images; the sample EMSE is 0.0764, which is approximately equal to the Wiener estimator's EMSE performance for this task.

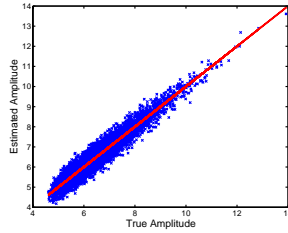


Fig. 14. Scatter plot of true versus scanning linear estimates of amplitude when the signal shape and location are known *a priori*. The red line indicates perfect estimation performance, and the actual estimates are blue points.

8. Conclusions

In this paper, we have described and evaluated two estimation procedures designed to minimize the EMSE under distinct statistical circumstances. The estimator's performance was evaluated for tasks of estimating signal location, amplitude, and radius from noisy data sets that include nuisance parameters. In our computation, the nuisance parameters were modeled using the Gaussian lumpy background model, and the signal was parameterized by a circle with a random centroid, radius, and amplitude.

The estimation task was modified to simulate various cases of prior knowledge. We measured the EMSE of the Wiener estimator when the radius was fixed, and the signal's location variability was decreased in steps approaching the limit of the location known exactly. Even for our simplified imaging model, the Wiener estimator did not perform well when the signal location was unknown. This result indicates a fundamental limitation on linear estimation. For instances when the location and size of the signal are both fixed, the Wiener estimator is a potentially useful linear estimation procedure for image-quality evaluation.

The scanning linear estimator was highly successful for the task of greatest difficulty, estimating signal size, amplitude, and location. In addition, when the location and radius of the signal are known and only signal amplitude is estimated, an analytic form that approximately minimizes the gradient in Eq. (59) can be used instead of searching for a global maximum. This proof-of-concept result will motivate further studies using this estimator.

The metric to be optimized (see Eq. (55)) by the scanning linear estimator offers an interesting connection to detection tasks. The first term of the objective function, derived from an approximate likelihood, is also known as the Hotelling discriminant function. In a detection study, the scalar value of the Hotelling discriminant is used as a threshold to classify signal-absent from signal-present images [1]. Instead of thresholding the output of the Hotelling discriminant function we have used it as an approximate likelihood to be maximized.

Acknowledgments

This work has been supported by NIH grants R37 EB000803 and P41 EB002035.