Noise histogram regularization for iterative image reconstruction algorithms

Samuel T. Thurman and James R. Fienup

The Institute of Optics, University of Rochester, Rochester, New York 14627, USA

Received May 22, 2006; revised September 21, 2006; accepted October 9, 2006;
posted October 13, 2006 (Doc. ID 71162); published February 14, 2007

We derive a regularization term for iterative image reconstruction algorithms based on the histogram of the residual difference between a forward-model image of a given object estimate and noisy image data. The term can be used to constrain this residual histogram to be statistically equivalent to the expected noise histogram, preventing overfitting of noise in a reconstruction. Reconstruction results from simulated imagery are presented for the cases of Gaussian and quantization noise.

OCIS codes: 100.3010, 100.3020.

1. INTRODUCTION

The typical image reconstruction problem can be stated as follows. Let $m \in \{1,2,\ldots,M\}$ and $n \in \{1,2,\ldots,N\}$ be integer sample indices on a two-dimensional Cartesian grid. For a space-invariant optical system, the noisy image $g(m,n)$ of a spatially incoherent object $f(m,n)$ can be modeled as

$$g(m,n) = s(m,n) * f(m,n) + \epsilon(m,n), \tag{1}$$

where $s(m,n)$ is the point-spread function (PSF) of the optical system, the * symbol represents a discrete two-dimensional convolution, and $\epsilon(m,n)$ represents a statistically independent realization of additive noise at each sample point. Given $g(m,n)$ and $s(m,n)$, and statistical knowledge of $\epsilon(m,n)$, we wish to reconstruct an object estimate $\hat{f}(m,n)$.

It is well known that the image reconstruction problem is ill posed, making regularization necessary to prevent overfitting of noise by many reconstruction algorithms. Here, we consider iterative reconstruction algorithms that use a nonlinear optimization routine, e.g., conjugate gradient, to find a reconstruction $\hat{f}(m,n)$ that maximizes or minimizes some metric, e.g., a likelihood metric. One regularization approach is to stop iterating when some convergence criteria are met.[1] Another approach is to include a term in the metric that prevents overfitting of noise by incorporating *a priori* information about either $f(m,n)$ or $\epsilon(m,n)$. The maximum entropy metric[2] incorporates knowledge that $f(m,n)$ must be nonnegative and the average over all possible objects is uniformly gray. The maximum residual likelihood metric[3] incorporates knowledge that $\epsilon(m,n)$ has no spatial structure.

Intuitively, the quality of $\hat{f}(m,n)$ is expected to increase as more *a priori* information about $f(m,n)$ and/or $\epsilon(m,n)$ is included in the reconstruction algorithm. Here, we derive a regularization term that incorporates *a priori* knowledge of the histogram of $\hat{\chi}(m,n)$, the normalized residual difference between a forward-model image of an object estimate and noisy image data given by

$$\hat{\chi}(m,n) = \frac{s(m,n) * \hat{f}(m,n) - g(m,n)}{\sigma(m,n)}, \tag{2}$$

where $\sigma(m,n)$ is the standard deviation of $\epsilon(m,n)$ at sample (m,n). The regularization term provides a means of quantifying the difference between the $\hat{\chi}(m,n)$ histogram and one expected from the statistics of $\chi(m,n) = \epsilon(m,n)/\sigma(m,n)$. Regularization can be achieved by using this noise histogram (NH) term to constrain the $\hat{\chi}(m,n)$ histogram to be statistically equivalent to one expected from the noise statistics. This NH term differs from the "exact error fitting" statistic of Bryan and Skilling,[4] which also incorporates *a priori* knowledge of the residual histogram, in that the NH term is based on an orthonormal expansion of the probability density function (PDF) of $\chi(m,n)$, whereas Bryan and Skilling's metric is based on the statistics of the sorted residuals. Also, Bryan and Skilling formulated their metric for Gaussian noise, whereas the NH term is more generally applicable to other types of noise.

Section 2 contains a review of PDF estimation using orthogonal expansions, upon which the NH term is based. Section 3 gives the formulation of the NH regularization term. Section 4 contains reconstruction results from simulated imagery for the cases of Gaussian and quantization noise. Conclusions are given in Section 5.

2. PROBABILITY DENSITY FUNCTION ESTIMATION

This section is a review of the technique of PDF estimation using orthonormal expansions from Refs. 5 and 6 that is needed for the derivation of the NH term in Section 3. Given J independent realizations of a random variable, x_j for $j \in \{1,2,\ldots,J\}$, with an unknown PDF $p(x)$, we wish to form an estimate $\hat{p}(x)$ of the PDF. When applied to the image reconstruction problem, $\{x_j\}$ represents the normalized residual $\hat{\chi}(m,n)$. For a set of appropriately chosen basis functions, $\psi_k(x)$ for $k \in \{1,2,\ldots,K\}$, the true PDF can be expressed as the sum

$$p(x) = \sum_{k=1}^{K} c_k \psi_k(x), \tag{3}$$

where c_k are expansion coefficients and $\psi_k(x)$ are basis functions that satisfy the orthonormality condition

$$\int_{-\infty}^{\infty} \psi_k(x)\psi_l(x)\mathrm{d}x = \delta_{k,l}, \tag{4}$$

where $\delta_{k,l}$ is the Kronecker delta function. The coefficients c_k are given by

$$c_k = \int_{-\infty}^{\infty} \psi_k(x)p(x)\mathrm{d}x = \langle \psi_k(x) \rangle, \tag{5}$$

where the angle brackets indicate an expectation value. Suppose we wish to find a PDF estimate of the same form as Eq. (3), i.e.,

$$\hat{p}(x) = \sum_{k=1}^{K} \hat{c}_k \psi_k(x), \tag{6}$$

where $\hat{c}_k$ are the expansion coefficients for $\hat{p}(x)$. Equation (5) suggests that the coefficients $\hat{c}_k$ can be estimated from $\{x_j\}$ using the sample-mean formula, i.e.,

$$\hat{c}_k = \frac{1}{J}\sum_{j=1}^{J} \psi_k(x_j). \tag{7}$$

Using well known results from statistics, coefficients $\hat{c}_k$ calculated in this manner have mean values μ_k and variances σ_k^2 given by

$$\mu_k \equiv \langle \hat{c}_k \rangle = c_k, \tag{8}$$

$$\sigma_k^2 \equiv \langle \hat{c}_k^2 \rangle - \langle \hat{c}_k \rangle^2 = \frac{\langle c_k^2 \rangle - c_k^2}{J}. \tag{9}$$

The PDF for each $\hat{c}_k$ is Gaussian with mean value c_k and standard deviation σ_k by the central limit theorem, regardless of the form of $p(x)$ (assuming both c_k and σ_k exist and are finite). Finally, the equations in this section are valid for a general $p(x)$.

3. REGULARIZATION TERM

In this section we use the results of Section 2 to formulate the NH regularization term. Suppose that the PDF of $\epsilon(m,n)$, the noise in an image at each pixel, has the same distribution with the exception of a scaling of the standard deviation $\sigma(m,n)$. Then the normalized noise $\chi(m,n)$ is equivalent to MN independent realizations of a unit-variance random variable having a PDF $p(x)$. By inspection of Eqs. (1) and (2), the normalized residual $\hat{\chi}(m,n)$ can be viewed as an estimate of $\chi(m,n)$. As such, the histogram of both quantities should be statistically equivalent. If $p(x)$ is known, and we have a suitable set of basis functions $\psi_k(x)$, then the results of Section 2 can be used to formulate an inequality that is satisfied when the two histograms are statistically equivalent as

$$\left| \frac{\hat{c}_k - c_k}{\sigma_k} \right| \le \kappa, \quad \forall\, k \in [1,2,\ldots,K], \tag{10}$$

where each $\hat{c}_k$ is calculated as

$$\hat{c}_k = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} \psi_k[\hat{\chi}(m,n)], \tag{11}$$

c_k and σ_k are calculated numerically via Eqs. (5) and (9), respectively, using $J=MN$, and κ is a parameter. In practice, κ is chosen to be the absolute value of the furthest outlier that would be expected for a set of K independent realizations of a normally distributed random variable. For example, $\kappa=4$ is a conservative value for $K=50$, since the probability of Eq. (10) not being satisfied is given by

$$1 - \left[\int_{-\kappa}^{\kappa} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2} \right) \mathrm{d}x \right]^K = 1 - \left[\mathrm{erf}\left(\frac{\kappa}{\sqrt{2}} \right) \right]^K = 0.32\,\%, \tag{12}$$

where erf(x) is the error function.

The NH regularization term $\Phi_{\mathrm{NH}}(\kappa)$ is formulated from relations (10) as a penalty function[7] of the form

$$\Phi_{\mathrm{NH}}(\kappa) = \sum_{k=1}^{K} \phi\left(\frac{\hat{c}_k - c_k}{\kappa\sigma_k} \right), \tag{13}$$

where $\phi(x)$ is a smooth continuous function defined as

$$\phi(x) = \begin{cases} 0 & \text{for } |x| \le 1 \\ \dfrac{(|x|-1)^2}{2} & \text{for } 1 < |x| \le 2 \\ |x| - 3/2 & \text{for } |x| > 2 \end{cases}. \tag{14}$$

Thus, $\Phi_{\mathrm{NH}}(\kappa)=0$ when Eqs. (10) are satisfied, and $\Phi_{\mathrm{NH}}(\kappa)$ grows increasingly large as these equations are violated. We use this particular form of $\phi(x)$ based on personal preference for mixed L1–L2 metrics over plain L2 metrics, but other forms of $\phi(x)$ are acceptable. In the case where a reconstruction $\hat{f}(m,n)$ is sought by minimizing some metric, regularization is achieved by adding $\Phi_{\mathrm{NH}}(\kappa)$ to the metric.

4. SIMULATION EXAMPLES

This section compares reconstruction results from simulated data using: (i) a Wiener-Helstrom filter[8,9] and an iterative reconstruction algorithm (ii) unregularized, (iii) regularized by stopping, and (iv) regularized by the NH term. The Wiener–Helstom reconstruction is given by

$$\hat{f}(m,n) = w(m,n) * g(m,n), \tag{15}$$

where $w(m,n)$ is the Wiener-Helstrom filter kernel given by the inverse discrete Fourier transform (DFT) of

$$W(u,v) = \frac{S^*(u,v)}{|S(u,v)|^2 + C\dfrac{\Phi_{\text{noise}}(u,v)}{\Phi_{\text{object}}(u,v)}}, \tag{16}$$

where (u,v) are spatial frequency sample indices, $S(u,v)$ is the DFT of $s(m,n)$, C is a parameter that can be adjusted to trade off image sharpness and noise suppression, and $\Phi_{noise}(u,v)$ and $\Phi_{object}(u,v)$ are the power spectra of $\epsilon(m,n)$ and $f(m,n)$, respectively. In our implementation,[9] $\Phi_{noise}(u,v)=\Phi_{noise}$ is assumed to be a constant and

$$\Phi_{object}(u,v) = \begin{cases} \sqrt{2}A^2 & \text{for } \rho(u,v) = 0 \\ A^2\rho^{-2\alpha} & \text{for } \rho(u,v) \neq 0 \end{cases}, \tag{17}$$

where $\rho(u,v)$ is the radial spatial frequency coordinate. The parameters Φ_{noise}, A, and α are determined from the DFT of the noisy image $G(u,v)$ by minimizing the following cost function:

$$e = \sum_{\substack{u=1 \\ \rho(u,v)\neq 0}}^{M} \sum_{v=1}^{N} \frac{1}{\rho(u,v)} \left\{ \frac{|G(u,v)|^2}{|S(u,v)|^2\Phi_{object}(u,v) + \Phi_{noise}} + \ln[|S(u,v)|^2\Phi_{object}(u,v) + \Phi_{noise}] \right\}. \tag{18}$$

The iterative algorithm uses a conjugate-gradient routine to find a $\hat{f}(m,n)$ that minimizes the normalized mean-square error metric Φ_{NMSE} given by

$$\Phi_{\text{NMSE}} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\hat{\chi}^2(m,n). \tag{19}$$

Since $\chi(m,n)$ is normalized to have unit variance, a reconstruction that is consistent with the noise statistics will result in $\Phi_{\text{NMSE}} \approx 1$. This suggests a regularization scheme based on stopping the iterations as soon as $\Phi_{\text{NMSE}} < 1$. For regularization using the NH term, we form a new metric $\Phi(\kappa,w)$ defined as

$$\Phi(\kappa,\lambda) = \Phi_{\text{NMSE}} + \lambda\Phi_{\text{NH}}(\kappa), \tag{20}$$

where λ is a weighting parameter. Following the standard implementation of penalty functions,[7] one would start an algorithm with a small λ and gradually increase λ until the constraints are satisfied (within some tolerable error). In the case where the metric and penalty function are both convex (with respect to the search parameters), it may be possible to simply use a predetermined, fixed value for λ. However, it is easy to show that $\Phi_{\text{NH}}(\kappa)$ is nonconvex with respect to $\hat{\chi}(m,n)$. Thus, we take the approach of increasing λ as iterations progress. In general, the starting value for λ, the prescription for increasing λ, and the optimization routine employed affect the convergence of an algorithm and therefore have a role in the regularization of $\hat{f}(m,n)$. We suggest starting with a λ sufficiently small to allow the algorithm to find a $\hat{f}(m,n)$ that is reasonably close to a $\hat{f}(m,n)$ that would be obtained without the penalty function. In the case where the penalty function is nonconvex, starting with too large a λ may result in a $\hat{f}(m,n)$ that satisfies the constraints, but is far from a minimum of the unregularized metric. In the case where the unregularized metric is convex, this ensures that the final $\hat{f}(m,n)$ will be fairly close to the global minimum of the unregularized metric. While Ref. 7 suggests increasing λ by approximately a factor of 10 after the optimization algorithm has met convergence criteria for the previous value of λ, we take the approach of increasing λ by a more modest factor, say 2, after running a fixed number of iterations, say 10 or 15, for the previous value of λ. This approach reduces the number of iterations per λ by eliminating the need to meet convergence criteria for each λ, which typically requires a significant number of iterations that result in very little progress through the solution space. However, this approach requires a larger number of λ values to obtain a $\hat{f}(m,n)$ that satisfies the constraints, since λ is increased by a more modest factor. It should be noted that increasing λ by too large a factor may cause the algorithm to stray off in the solution space, away from any minimum of the unregularized metric that we tried to approach by starting with a sufficiently small λ. An analytic expression for the gradient of $\Phi(\kappa,w)$ with respect to the $\hat{f}(m,n)$ sample values is given in Appendix

(a)

(b)

Fig. 1. Object $f(m,n)$ used for simulations: (a) whole 256×256 object and (b) zoomed subsection shown for comparison with reconstruction results.

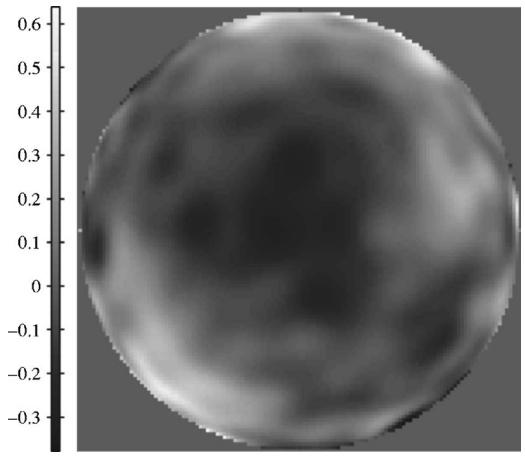

Fig. 2. Pupil phase error in units of waves.

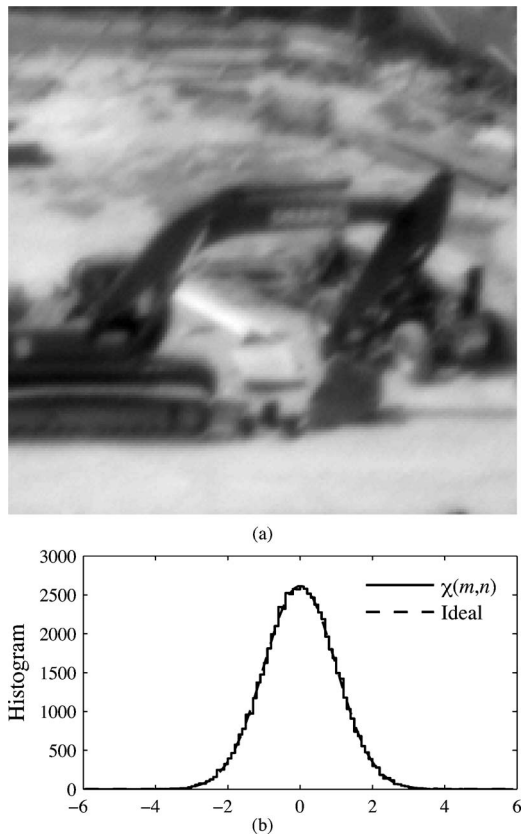

Fig. 3. (a) Noisy image $g(m,n)$ of object shown in Fig. 2 with Gaussian additive noise with $\sigma=2$ (compared with mean image value of 50), and (b) the histogram of $\chi(m,n)$ and the ideal histogram of a normally distributed random variable.

A. The simulations used the 256×256 object shown in Fig. 1(a) for $f(m,n)$, and a Nyquist-sampled PSF, $s(m,n)$, for an optical system with a circular aperture having the phase error shown in Fig. 2, which is a simulated phase screen[10] for atmospheric turbulence with $D/r_0=2$. The reconstruction results are compared numerically using the resulting values of Φ_{NMSE}, $\Phi_{\mathrm{NH}}(\kappa)$, and the minimum normalized root-mean-square (NRMS) error E defined by Eq. (13) of Ref. 11, which measures the difference between $\hat{f}(m,n)$ and $f(m,n)$. The following subsections contain reconstruction results for cases of Gaussian and quantization noise.

A. Gaussian Noise

Figure 3(a) shows the simulated noisy image $g(m,n)$, which includes Gaussian noise having a standard deviation $\sigma=1/\sqrt{12}$ as compared with a mean value of 50 for $s(m,n)*f(m,n)$. This σ was chosen to yield the same signal-to-noise ratio as the quantization noise case considered in Subsection 4.B. Figure 3(b) shows the histogram of $\chi(m,n)$ in comparison with the ideal histogram. For Gaussian noise, the PDF of $\chi(m,n)$ is

$$p(x)=1/\sqrt{2\pi}\exp(-x^2/2). \tag{21}$$

As such, a natural choice for the orthogonal functions $\psi_k(x)$ are the Hermite–Gauss functions, i.e.,

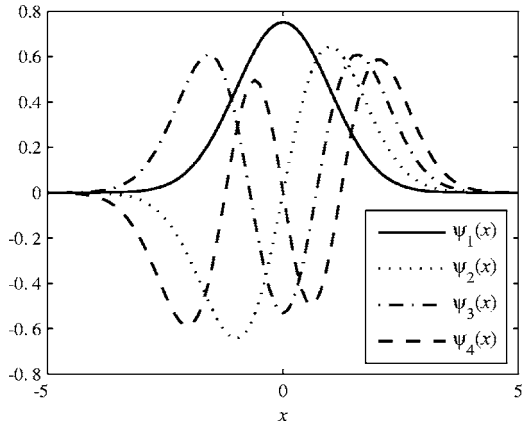

Fig. 4. Plot of the first four basis functions $\psi_k(x)$.

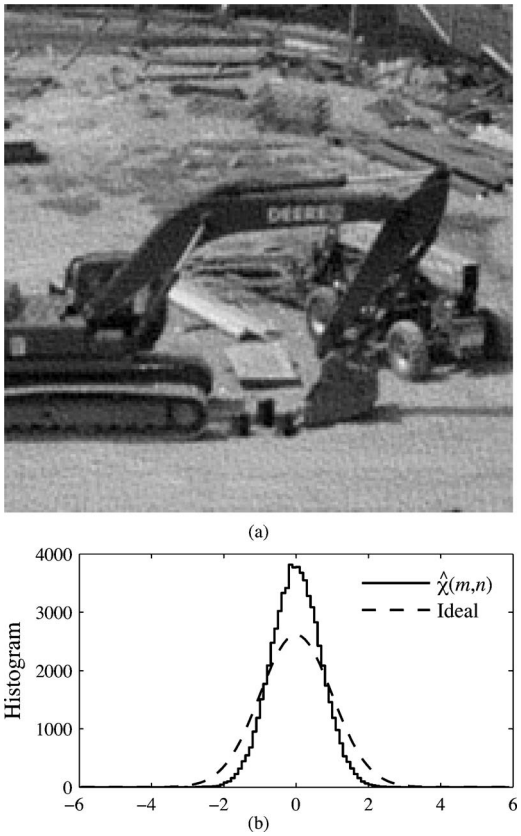

Fig. 5. Wiener–Helstrom reconstruction for Gaussian noise: (a) $\hat{f}(m,n)$, (b) histogram of $\hat{\chi}(m,n)$.

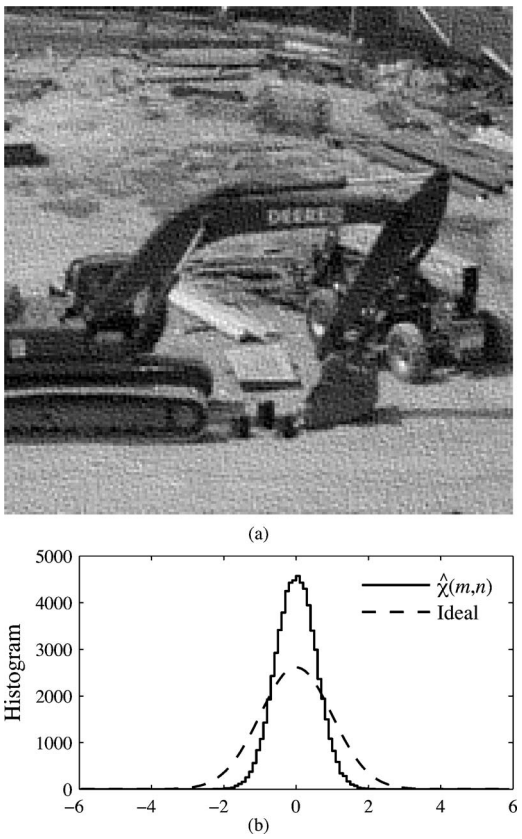

Fig. 6. Iterative reconstruction for Gaussian noise, unregularized: (a) $\hat{f}(m,n)$, (b) histogram of $\hat{\chi}(m,n)$.

$$\psi_k(x) = \alpha_k H_{k-1}(x)\exp(-x^2/2), \tag{22}$$

where $H_k(x)$ is the kth order Hermite polynomial as defined in Ref. 12 and α_k is a normalization constant. The first four $\psi_k(x)$ are shown in Fig. 4. For the case of regularization by the NH term, $\Phi_{\mathrm{NH}}(\kappa)$ was calculated using $K=25$ basis functions with $\kappa=4$. Iterations started with a weighting parameter value of $\lambda=10^{-4}$, which was increased by a factor of 2 after every 10 iterations. This approach of steadily increasing λ allows the algorithm to quickly find a reconstruction with a small Φ_{NMSE} value while gradually increasing the effect of $\Phi_{\mathrm{NH}}(\kappa)$ until the constraints in Eq. (10) are adequately satisfied. The starting value of λ and the number of iterations per λ were chosen to allow the algorithm to obtain a $\Phi_{\mathrm{NMSE}}\approx 1$ within the first 2–3 values of λ.

Figures 5–8 show the various reconstructions and the corresponding histograms of $\hat{\chi}(m,n)$, while Table 1 lists the total number of iterations and the numerical values of Φ_{NMSE}, $\Phi_{\mathrm{NH}}(\kappa)$, and E for the true object [calculated using $\hat{f}(m,n)=f(m,n)$], $g(m,n)$, and each reconstruction $\hat{f}(m,n)$. Each reconstruction is displayed on the same gray scale as Fig. 1(b). The Wiener–Helstrom reconstruction, shown in Fig. 5(a), exhibits some high-spatial-frequency correlated noise. This is because the Wiener-Helstrom reconstruction is formed by boosting the spatial-frequency components of the noisy image to compensate for the optical transfer function of the system, while taking into account the signal-to-noise ratio in the spatial-frequency domain. In this process, higher spatial frequencies are boosted more than low spatial frequencies, yielding high-spatial-frequency colored noise in the reconstruction. That the Wiener–Helstrom reconstruction is inconsistent with the noise statistics is evident from Fig. 5(b), which shows that the $\hat{\chi}(m,n)$ histogram is narrower than expected and by data from Table 1, which indicate that $\Phi_{\mathrm{NMSE}}=0.4783$ is less than unity and $\Phi_{\mathrm{NH}}(\kappa)=57.28$ is relatively large.

A value of the Wiener–Helstrom filter parameter of $C=1$ typically yields the lowest reconstruction error E, but E alone is not the best indicator of overall image quality. C can be increased to reduce the amplitude of the colored noise at the expense of the image sharpness, or C can be reduced to increase both image sharpness and the noise amplitude. While Ref. 9 suggests that image analysts prefer a value of $C=0.2$ on average, we chose a value of 0.5 as a tradeoff between image sharpness and noise amplitude.

The unregularized iterative reconstruction shown in Fig. 6(a) exhibits excessive amounts of high-spatial-frequency noise because the algorithm tends to overfit the noise, which results in a small value of $\Phi_{\mathrm{NMSE}}=0.3149$, large values of $\Phi_{\mathrm{NH}}(\kappa)=105.9$ and $E=0.1092$, and a $\hat{\chi}(m,n)$ histogram that is much narrower than expected. When regularized by stopping as soon as $\Phi_{\mathrm{NMSE}}<1$, the reconstruction, shown in Fig. 7(a), yields numerical values of $\Phi_{\mathrm{NMSE}}=0.9980$ and $\Phi_{\mathrm{NH}}(\kappa)=2.078$, and a $\hat{\chi}(m,n)$ histogram that are all consistent with the noise statistics, as well as a reasonably low reconstruction error of $E=0.0748$, but the reconstruction is much less sharp than any of the other reconstructions. Stopping the algorithm when $\Phi_{\mathrm{NMSE}}\approx 1$ ensures that the resulting $\hat{\chi}(m,n)$ will

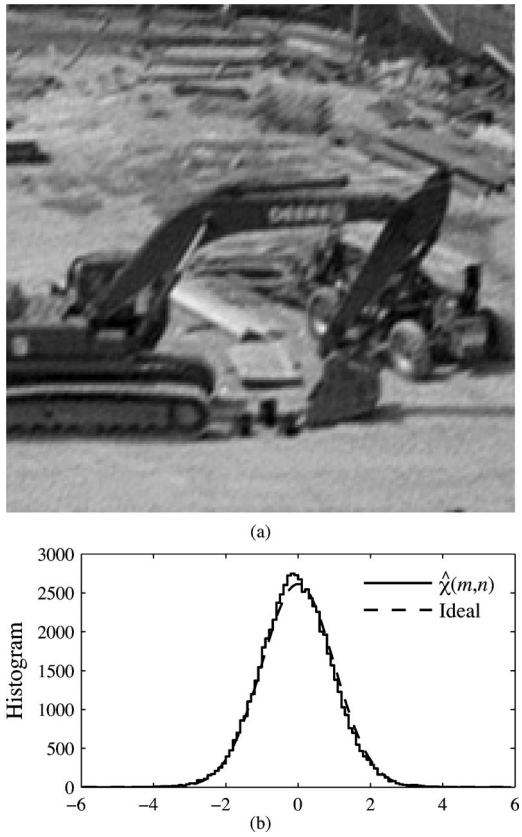

Fig. 7. Iterative reconstruction for Gaussian noise, regularized by stopping when $\Phi_{\mathrm{NMSE}}<1$: (a) $\hat{f}(m,n)$, (b) histogram of $\hat{\chi}(m,n)$.

have the appropriate variance, but does not ensure that $\hat{\chi}(m,n)$ will have the appropriate distribution. For reasons we do not completely understand, the regularization imposed by the conjugate gradient minimization routine tends to yield a $\hat{\chi}(m,n)$ with a Gaussian distribution (see also the reconstruction results for quantization noise).

Additionally, the use of conjugate gradient with Φ_{NMSE} tends to converge on the low spatial frequencies of $\hat{f}(m,n)$ rather quickly, but requires more iterations to converge on the high spatial frequencies. Thus, while $\hat{\chi}(m,n)$ is consistent with the expected noise histogram, the reconstruction is not very sharp. Figure 8(a) shows the iterative reconstruction regularized by $\Phi_{\mathrm{NH}}(\kappa)$, for which $\Phi_{\mathrm{NMSE}}=0.9190$, $\Phi_{\mathrm{NH}}(\kappa)=0.1369$, the $\hat{\chi}(m,n)$ histogram are consistent with the noise statistics, and the reconstruction error $E=0.0683$ is low. Comparing the Wiener–Helstrom and NH regularized iterative reconstructions closely, with reference to $f(m,n)$ in Fig. 1, the NH-regularized reconstruction appears slightly less sharp, but the amplitude of the colored noise appears considerably smaller.

Figure 9 shows plots of the values of Φ_{NMSE}, $\Phi_{\mathrm{NH}}(\kappa)$, and E as a function of the iteration number for the various reconstruction algorithms. Referring to Fig. 9(a), notice that, without the $\Phi_{\mathrm{NH}}(\kappa)$ regularization term, Φ_{NMSE} decreases monotonically with each iteration, but with the regularization term Φ_{NMSE} eventually settles on a value near unity, which is consistent with the noise statistics. Referring to Fig. 9(b), notice that E decreases rapidly with the first few iterations. Without the NH regularization term, E reaches a minimum value at an iteration when $\Phi_{\mathrm{NMSE}}\approx 1$ (note that Table 1 indicates that Φ_{NMSE} decreases below unity on iteration 17), but then increases as the algorithm continues to fit noise with each iteration. With the NH regularization term, E does not increase appreciably after reaching a minimum value. The Wiener–Helstrom filter has a computational advantage in that it achieves a reconstruction error E equivalent to the mini-

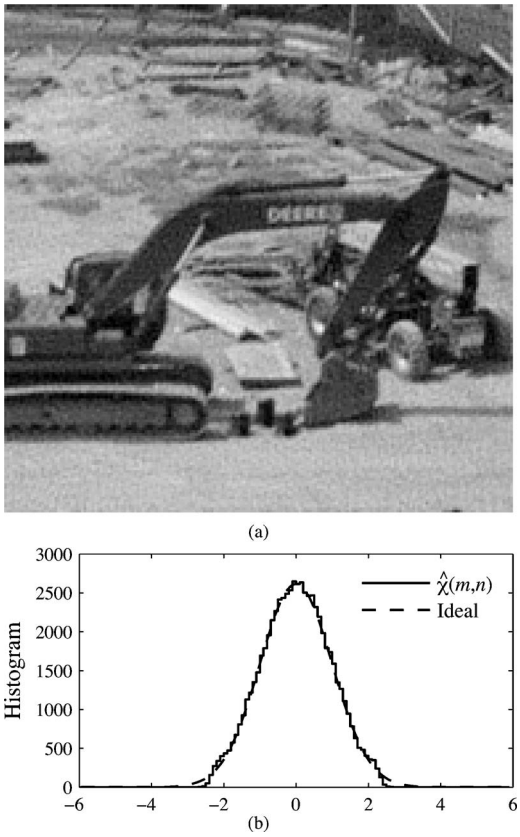

Fig. 8. Iterative reconstruction for Gaussian noise, regularized by Φ_{NH}: (a) $\hat{f}(m,n)$, (b) histogram of $\hat{\chi}(m,n)$.

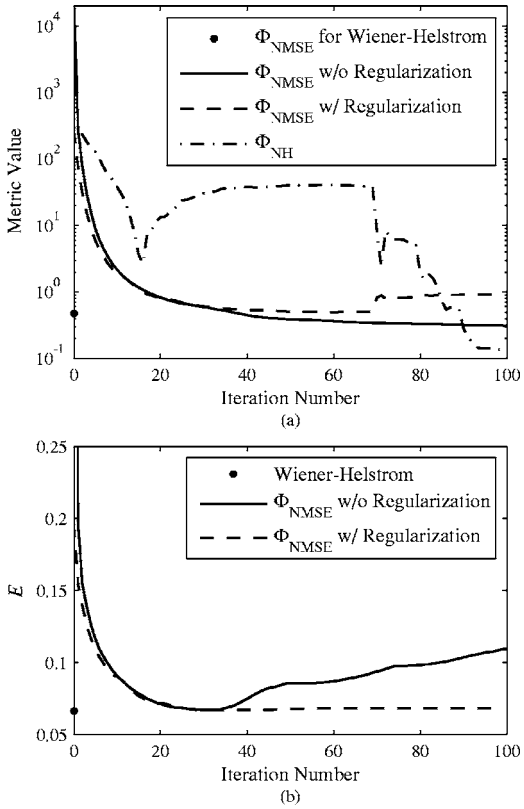

Fig. 9. Minimum NRMS error in each reconstruction statistics as a function of iteration number for Gaussian noise.

Table 1. Reconstruction Results for Gaussian Noise

Algorithm	Regularization	Iterations	$\Phi_{\mathrm{NMSE}}{}^{a}$	$\Phi_{\mathrm{NH}}(\kappa)^{a}$	E^{a}
True object	None	0	0.9981	0	0
Noisy Image	None	0	106.3	231.0	0.1547
Wiener–Helstrom	Filter parameter $C=0.5$	0	0.4783	57.28	0.0666
Φ_{NMSE}	None	100	0.3149	105.9	0.1092
Φ_{NMSE}	Stopping when $\Phi_{\mathrm{NMSE}}<1$	17	0.9980	2.078	0.0748
Φ_{NMSE}	$\lambda\Phi_{\mathrm{NH}}(\kappa)$	100	0.9190	0.1369	0.0683

[a]Ideal values for the metrics Φ_{NMSE} and $\Phi_{\mathrm{NH}}(\kappa)$ and the reconstruction error E are $\Phi_{\mathrm{NMSE}}=1$, $\Phi_{\mathrm{NH}}(\kappa)=0$, and $E=0$.

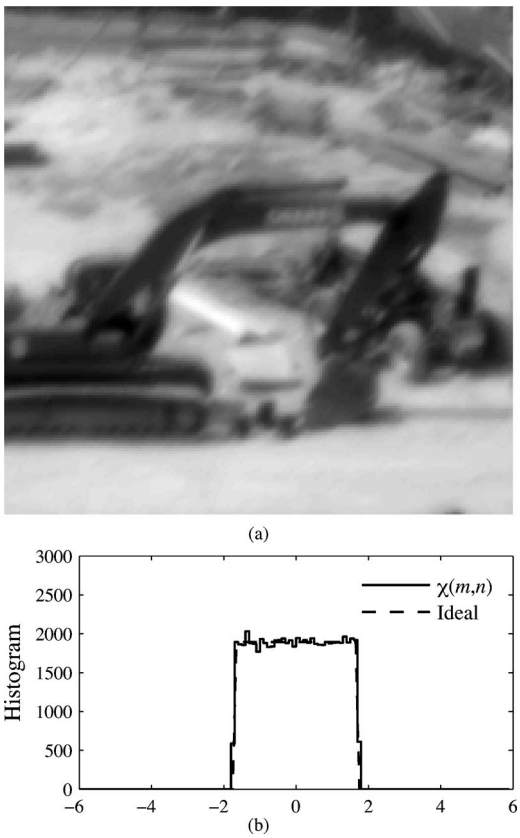

Fig. 10. (a) Noisy image $g(m,n)$ of object shown in Fig. 2 quantized to integer values ($\sigma=1/\sqrt{12}$), (b) histogram of $\chi(m,n)$ and the ideal histogram for a uniform distribution.

mum value for the iterative algorithms with no iterations. Alternatively, an iterative algorithm has the ability to include additional information about the noise statistics or object and accurately model nonlinear aspects of an imaging system.

B. Quantization Noise

The quantization noise case is included to highlight certain aspects of the $\Phi_{\mathrm{NH}}(\kappa)$ regularization term. To create an image with quantization noise, a noiseless image of $f(m,n)$ was rounded to integer values. Figure 10 shows the noisy quantized image along with a histogram of $\chi(m,n)$, where $\sigma=1/\sqrt{12}$. For quantization noise, the PDF of $\chi(m,n)$ is

$$p(x)=\begin{cases}\dfrac{1}{\sqrt{12}} & \text{for } |x|\le\dfrac{\sqrt{12}}{2}\\[2ex] 0 & \text{for } |x|>\dfrac{\sqrt{12}}{2}\end{cases}. \tag{23}$$

Although the functions in Eq. (22) may not be the most efficient for expanding $p(x)$ in a series of the form of Eq. (3), they can be used to evaluate the NH regularization term. Figures 11–14 show reconstruction results using each algorithm, Table 2 lists numerical details for the true object and each reconstruction, and Fig. 15 is a graph of the metric values and reconstruction error versus iteration number for each algorithm. The iterative algorithm regularized by the NH term (using $K=25$ basis functions and $\kappa=4$) started with a weighting parameter value of $\lambda=10^{-4}$, which was increased by a factor of 2 after every 15 iterations. While these reconstruction results are very similar to the reconstruction results obtained for Gaussian noise, there are some important differences. In Figs. 11–13 notice that each of the $\hat{\chi}(m,n)$ histograms basically has a Gaussian shape. The histogram in Fig. 13(b) is especially interesting, because regularization of the iterative algorithm by stopping the iterations as soon as $\Phi_{\mathrm{NMSE}}<1$ results in a $\hat{\chi}(m,n)$ with the appropriate variance having a Gaussian-like distribution, which appears to be the result of regularization imposed by the minimization routine, as mentioned earlier. In comparison, notice that the $\hat{\chi}(m,n)$ histogram for the NH-regularized reconstruction, shown in Fig. 14(b), agrees well with the ideal histogram expected from the noise statistics. This is a result of the fact that $\Phi_{\mathrm{NH}}(\kappa)$ incorporates information about the higher-order moments of $\hat{\chi}(m,n)$ through the $\psi_k(x)$ functions and the c_k coefficients. The Gibb's ringing artifacts in the $\hat{\chi}(m,n)$ in Fig. 14(b) are due to the fact that the discontinuous PDF in Eq. (23) cannot be completely described by an expansion using the continuous functions in Eq. (22).

5. CONCLUSIONS

We have derived a regularization term for image reconstruction that penalizes reconstructions for which the residual histogram is statistically inconsistent with that expected from the noise statistics. Simulation results

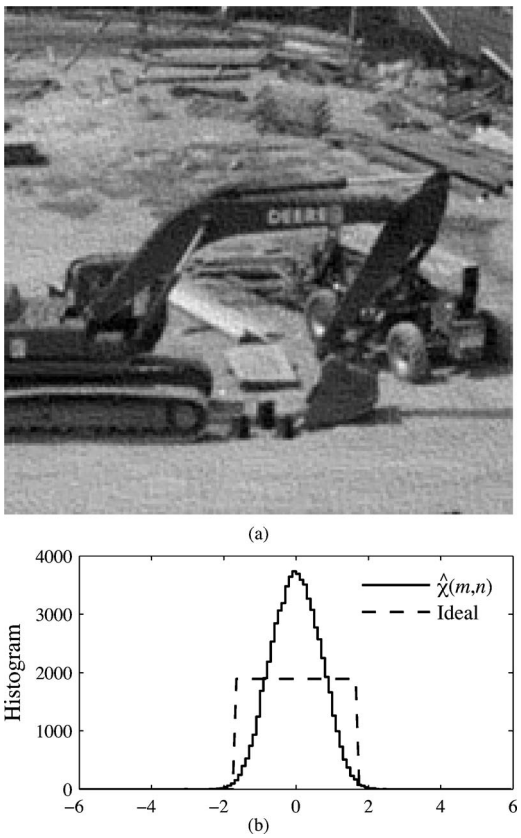

Fig. 11. Wiener–Helstrom reconstruction for quantization noise: (a) $\hat{f}(m,n)$, (b) histogram of $\hat{\chi}(m,n)$.

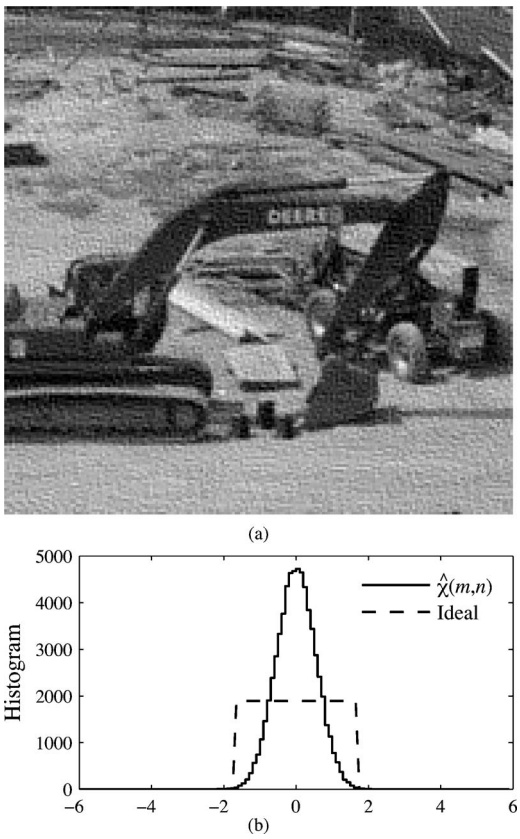

Fig. 12. Iterative reconstruction for quantization noise, unregularized: (a) $\hat{f}(m,n)$, (b) histogram of $\hat{\chi}(m,n)$.

demonstrate regularization for both Gaussian and quantization noise cases. In general, the regularization term can accommodate any zero-mean noise source for which the noise PDF at each image pixel is the same except for a scaling of the standard deviation. Thus, the method can handle the case of photon noise only in the limit that the Possion distribution for the noise at each pixel can be modeled as a Gaussian distribution, which is a reasonable approximation when the minimum single-pixel photon count for an image is on the order of 20 or more photons (see Appendix A). This approach avoids the problems with nonregularized algorithms of overfitting the noise and not knowing when to stop the iterations. While we only discussed image reconstruction with a known system PSF, the regularization term is directly applicable to the problem of jointly estimating the PSF and the object.

APPENDIX A

This appendix contains an analytic expression for the partial derivatives of the $\Phi(\kappa,\lambda)$ with respect to the sample values of a reconstruction $\hat{f}(m,n)$. First, the following definitions are used to simplify notation

$$\hat{g}(m,n) = s(m,n) * \hat{f}(m,n) = \sum_{m'=1}^{M}\sum_{n'=1}^{N} s(m-m',n-n')\hat{f}(m',n'), \quad \text{(A1)}$$

$$\psi_k'(x) = \frac{\mathrm{d}\psi_k(x)}{\mathrm{d}x}, \quad \text{(A2)}$$

$$\phi'(x) = \frac{\mathrm{d}\phi(x)}{\mathrm{d}x}. \quad \text{(A3)}$$

Differentiating Eq. (20), we can write

$$\frac{\partial\Phi(\kappa,\lambda)}{\partial\hat{f}(m,n)} = \frac{\partial\Phi_{\mathrm{NMSE}}}{\partial\hat{f}(m,n)} + \lambda\frac{\partial\Phi_{\mathrm{NH}}(\kappa)}{\partial\hat{f}(m,n)}. \quad \text{(A4)}$$

Through repeated application of the chain rule of differentiation and Eqs. (2), (11), (13), (19), and (A1)–(A3), one can show that

$$\begin{aligned}\frac{\partial\Phi_{\mathrm{NMSE}}}{\partial\hat{f}(m,n)} &= \frac{2}{MN}\sum_{m'=1}^{M}\sum_{n'=1}^{N}\hat{\chi}(m',n')\frac{\partial\hat{\chi}(m',n')}{\partial\hat{f}(m,n)}\\ &= \frac{2}{MN}\sum_{m'=1}^{M}\sum_{n'=1}^{N}\frac{\hat{\chi}(m',n')}{\sigma(m',n')}\frac{\partial\hat{g}(m',n')}{\partial\hat{f}(m,n)}\\ &= \frac{2}{MN}\sum_{m'=1}^{M}\sum_{n'=1}^{N}\frac{\hat{\chi}(m',n')}{\sigma(m',n')}s(m'-m,n'-n),\end{aligned} \quad \text{(A5)}$$

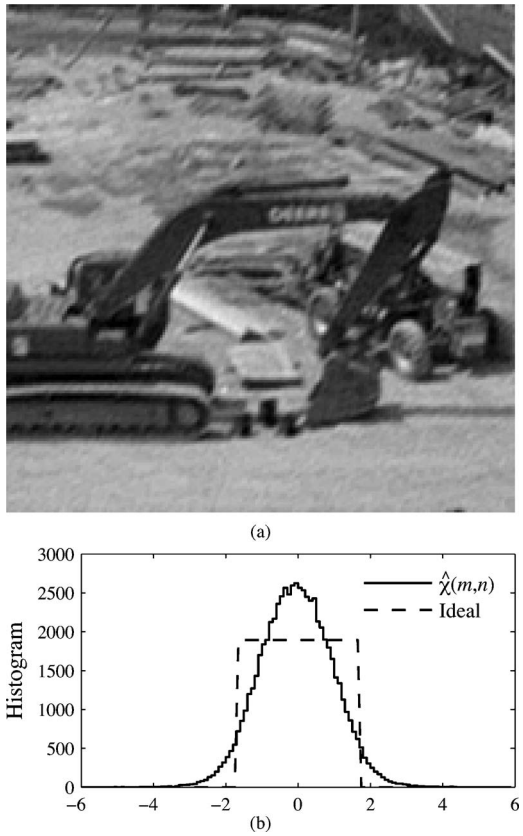

Fig. 13. Iterative reconstruction for quantization noise, regularized by stopping when $\Phi_{\mathrm{NMSE}}<1$: (a) $\hat{f}(m,n)$, (b) histogram of $\hat{\chi}(m,n)$.

$$\frac{\partial \Phi_{\mathrm{NH}}(\kappa)}{\partial \hat{f}(m,n)} = \sum_{k=1}^{K} \frac{1}{\kappa\sigma_k} \phi'\left(\frac{\hat{c}_k - c_k}{\kappa\sigma_k}\right) \frac{\partial \hat{c}_k}{\partial \hat{f}(m,n)}$$

$$= \frac{1}{MN} \sum_{m'=1}^{N} \sum_{n'=1}^{N} \sum_{k=1}^{K} \frac{1}{\kappa\sigma_k} \phi'\left(\frac{\hat{c}_k - c_k}{\kappa\sigma_k}\right) \times \psi_k'[\hat{\chi}(m',n')] \frac{\partial \hat{\chi}(m',n')}{\partial \hat{f}(m,n)}$$

$$= \frac{1}{MN} \sum_{m'=1}^{N} \sum_{n'=1}^{N} \sum_{k=1}^{K} \frac{1}{\kappa\sigma_k} \phi'\left(\frac{\hat{c}_k - c_k}{\kappa\sigma_k}\right) \times \psi_k'[\hat{\chi}(m',n')] \frac{1}{\sigma(m',n')} \frac{\partial \hat{g}(m',n')}{\partial \hat{f}(m,n)}$$

$$= \frac{1}{MN} \sum_{m'=1}^{N} \sum_{n'=1}^{N} \sum_{k=1}^{K} \frac{1}{\kappa\sigma_k} \phi'\left(\frac{\hat{c}_k - c_k}{\kappa\sigma_k}\right) \times \psi_k'[\hat{\chi}(m',n')] \frac{1}{\sigma(m',n')} s(m'-m,n'-n). \tag{A6}$$

For the case of photon noise, the noise standard deviation $\sigma(m,n)$ is signal dependent and can be estimated by

$$\hat{\sigma}(m,n) = \sqrt{\hat{g}(m,n)}, \tag{A7}$$

in which case

$$\hat{\chi}(m,n) = \frac{\hat{g}(m,n) - g(m,n)}{\hat{\sigma}(m,n)}, \tag{A8}$$

(a)

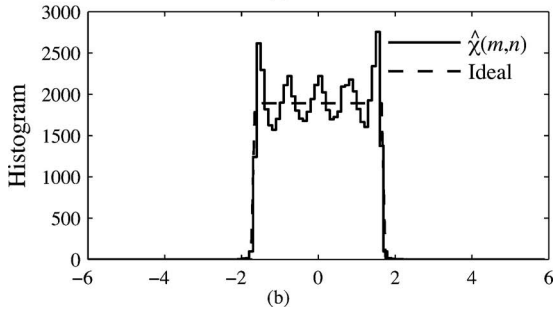

Fig. 14. Iterative reconstruction for quantization noise, regularized by Φ_{NH}: (a) $\hat{f}(m,n)$, (b) histogram of $\hat{\chi}(m,n)$.

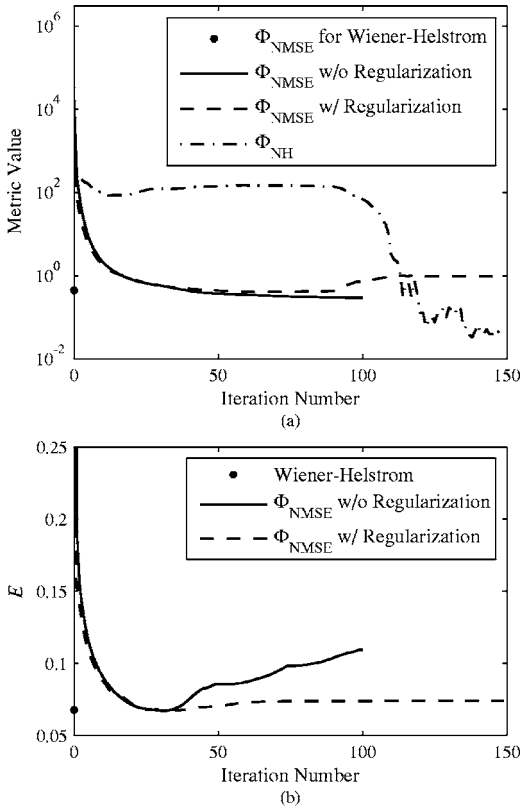

Fig. 15. Minimum NRMS error in each reconstruction statistics as a function of iteration number for quantization noise.

Table 2. Reconstruction Results for Quantization Noise

Algorithm	Regularization	Iterations	$\Phi_{\mathrm{NMSE}}{}^a$	$\Phi_{\mathrm{NH}}{}^a$	E^a
True Object	None	0	1.0014	0	0
Noisy Image	None	0	106.4	230.7	0.1547
Wiener–Helstrom	Filter parameter $C=0.5$	0	0.4487	168.1	0.0677
Φ_{NMSE}	None	100	0.2968	207.8	0.1093
Φ_{NMSE}	Stopping when $\Phi_{\mathrm{NMSE}}<1$	17	0.9841	98.08	0.0751
Φ_{NMSE}	$\lambda\Phi_{\mathrm{NH}}(\kappa)$	150	0.9761	0.0532	0.0743

[a]Ideal values for the metrics Φ_{NMSE} and $\Phi_{\mathrm{NH}}(\kappa)$ and the reconstruction error E are $\Phi_{\mathrm{NMSE}}=1$, $\Phi_{\mathrm{NH}}(\kappa)=0$, and $E=0$.

$$\frac{\partial \hat{\chi}(m,n)}{\partial \hat{g}(m,n)} = \frac{1}{\hat{\sigma}(m,n)}\left[1 - \frac{\hat{\chi}(m,n)}{2\hat{\sigma}(m,n)}\right], \tag{A9}$$

$$\frac{\partial \Phi_{\mathrm{NH}}(\kappa)}{\partial \hat{f}(m,n)} = \frac{1}{MN}\sum_{m'=1}^{N}\sum_{n'=1}^{N}\sum_{k=1}^{K}\frac{1}{\kappa\sigma_k}\phi'\left(\frac{\hat{c}_k - c_k}{\kappa\sigma_k}\right)\psi_k'[\hat{\chi}(m',n')] \times \frac{1}{\hat{\sigma}(m',n')}\left[1 - \frac{\hat{\chi}(m',n')}{2\hat{\sigma}(m',n')}\right]s(m'-m,n'-n), \tag{A10}$$

and

$$\frac{\partial \Phi_{\mathrm{NMSE}}}{\partial \hat{f}(m,n)} = \frac{2}{MN}\sum_{m'=1}^{M}\sum_{n'=1}^{N}\frac{\hat{\chi}(m',n')}{\hat{\sigma}(m,n)}\left[1 - \frac{\hat{\chi}(m,n)}{2\hat{\sigma}(m,n)}\right] \times s(m'-m,n'-n). \tag{A11}$$

ACKNOWLEDGMENTS

This work was supported by Lockheed Martin Corporation. Photo credits for the image used in the simulations go to Ryan T. DeRosa.

Corresponding author Samuel T. Thurman's e-mail address is thurman@optics.rochester.edu.

REFERENCES

1. H. J. Trussell, "Convergence criteria for iterative restoration methods," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-31**, 129–136 (1983).
2. B. R. Frieden, *Probability, Statistical Optics, and Data Testing: a Problem Solving Approach* (Springer, 2001), pp. 285–286.
3. R. K. Piña and R. C. Puetter, "Incorporation of spatial information in Bayesian image reconstruction: the maximum residual likelihood criterion," Publ. Astron. Soc. Pac. **104**, 1096–1103 (1992).
4. R. K. Bryan and J. Skilling, "Deconvolution by maximum entropy, as illustrated by application to the jet of M87," Mon. Not. R. Astron. Soc. **191**, 69–79 (1980).
5. N. N. Cencov, "Evaluation of an unknown distribution density from observations," Sov. Math. Dokl. **3**, 1559–1562 (1962).
6. B. R. Frieden, *Probability, Statistical Optics, and Data Testing: a Problem Solving Approach* (Springer, 2001), pp. 277–305.
7. D. M. Ryan, "Penalty and barrier functions," in *Numerical Methods for Constrained Optimization*, P. E. Gill and W. Murray, eds. (Academic, 1974), pp. 175–190.
8. C. W. Helstrom, "Image restoration by the method of least squares," J. Opt. Soc. Am. **57**, 297–303 (1967).
9. J. R. Fienup, D. Griffith, L. Harrington, A. M. Kowalczyk, J. J. Miller, and J. A. Mooney, "Comparison of reconstruction algorithms for images from sparse-aperture systems," in *Image Reconstruction from Incomplete Data II*, P. J. Bones, M. A. Fiddy, and R. P. Millane, eds., Proc. SPIE **4792**, 1–8 (2002).
10. N. Roddier, "Atmospheric wavefront simulation using Zernike polynomials," Opt. Eng. (Bellingham) **29**, 1174–1180 (1990).
11. J. R. Fienup, "Invariant error metrics for image reconstruction," Appl. Opt. **36**, 8352–8357 (1997).
12. G. E. Andrews, R. Askey, and R. Roy, *Special Functions* (Cambridge, 1999), Sec. 6.1, pp. 278–282.