Aberration correction by maximizing generalized sharpness metrics

J. R. Fienup

University of Rochester, Institute of Optics, 410 Wilmot Building, 275 Hutchison Road, Rochester,
New York 14627-0186

J. J. Miller

Veridian Systems Division, P.O. Box 134008, Ann Arbor, Michigan 48113-4008

The technique of maximizing sharpness metrics has been used to estimate and compensate for aberrations with adaptive optics, to correct phase errors in synthetic-aperture radar, and to restore images. The largest class of sharpness metrics is the sum over a nonlinear point transformation of the image intensity. How the second derivative of the point nonlinearity varies with image intensity determines the effects of various metrics on the imagery. Some metrics emphasize making shadows darker, and other emphasize making bright points brighter. One can determine the image content needed to pick the best metric by computing the statistics of the image autocorrelation or of the Fourier magnitude, either of which is independent of the phase error. Computationally efficient, closed-form expressions for the gradient make possible efficient search algorithms to maximize sharpness. © 2003 Optical Society of America

OCIS codes: 100.5070, 280.6730, 100.3010, 100.3020, 110.3000, 110.5100, 010.1080.

1. INTRODUCTION

When an imaging system experiences phase errors (aberrations), the impulse response of the system is blurred: the peak value decreases, the mainlobe width increases, and the sidelobes increase. The resultant blurred image has reduced sharpness and contrast. It has long been recognized that one can estimate the phase errors, and correct them, by finding the phase-error estimate that maximizes the sharpness of the image. This technique has been applied not only to incoherent optical imaging[1] but to coherent imaging with synthetic-aperture radar (SAR) as well.[2–4] There are many variations on what is meant by the sharpness of the image and how one goes about maximizing it. In this paper we restrict our attention to sharpness metrics that are given by the sum over a nonlinear point transformation of the image intensity. Four of Muller and Buffington's sharpness metrics[1] (numbers 1, 2, 5, and 7) fit this description. This paper explains how the actions of the various sharpness metrics can be understood by examining the form of the second derivative of the nonlinear point transformation. The knowledge thus obtained allows us to select better metrics depending on expected scene content and to invent "designer metrics."

Section 2 explains the generalized efficient sharpening algorithm in the context of phase-error correction for SAR. Section 3 shows some computer simulation results that indicate how different sharpness metrics behave for different types of images. Section 4 discusses the effect of the second derivative of the point nonlinearity and the phenomenology of SAR images that influences our choices for it. Section 5 compares results from SAR focusing

with different metrics, including some designer metrics. Section 6 presents conclusions.

2. SYNTHETIC-APERTURE-RADAR FOCUSING BY EFFICIENT SHARPENING ALGORITHM

Although image sharpening was originally conceived for incoherent imaging systems,[1] Paxman and Marron[2] showed that it can be applied to coherent imaging systems as well. We consider for simplicity spotlight-mode SAR,[5] for which polar formatting has already been performed. Then the signal-history data is related to the image by a Fourier transform. Each pulse of the data may be first Fourier transformed in the frequency dimension to arrive at the range-compressed signal history. The phase error is typically a function of the pulse number (or the along-track, or azimuth, or "slow-time" coordinate), so our model for the degraded range-compressed signal history is

$$G_d(x, \nu) = F(x, \nu)\exp[i\,\phi_e(\nu)], \tag{1}$$

where x is the range coordinate, ν is the slow-time coordinate, $F(x, \nu)$ is the ideal range-compressed signal history, and $\phi_e(\nu)$ is the one-dimensional (1-D) phase error. For very large phase errors or large angles of data collection, a more complicated model may be needed. For simplicity we consider imaging in the slant-range plane. The corresponding degraded image is given by inverse Fourier transforming in the azimuth dimension,

$$g_d(x, y) = \text{FT}^{-1}[G_d(x, \nu)]$$

$$= (1/N)\sum_x G_d(x, \nu)\exp(i2\pi\nu y/N), \quad (2)$$

where FT represents a 1-D discrete inverse Fourier transform, y is the image-domain azimuth (cross-range) coordinate, and N is the number of pixels in the azimuth dimension. A phase-error estimate, $\phi(\nu)$, is used to compute the corrected signal history,

$$G(x, \nu) = G_d(x, \nu)\exp[-i\phi(\nu)], \quad (3)$$

and the corresponding complex-valued image,

$$g(x, y) = \text{FT}^{-1}[G(x, \nu)]. \quad (4)$$

The image intensity is

$$I(x, y) = |g(x, y)|^2. \quad (5)$$

The image-sharpening approach is to iteratively choose $\phi(\nu)$ to maximize the sharpness of the image. We do this in an efficient manner as follows. We define a nonlinear point transformation on the image intensity, $\Gamma[I(x, y)]$, and compute a weighted sum over all the pixels to compute the sharpness metric,

$$S_\Gamma = \sum_{x,y} w(x)\Gamma[I(x, y)]. \quad (6)$$

For example, Muller and Buffington's fifth sharpness metric[1] involves raising the image to a power:

$$\Gamma[I(x, y)] = [I(x, y)]^\beta. \quad (7)$$

For SAR we may also include a function $w(x)$ that allows us to weight each range bin according to an estimate of how useful it is for phase-error estimation.[4] For example, if a range bin has more energy or lower variance than the average, it might be given more or less weight, respectively. The sharpness metric is a function of the phase-error estimate, so we can perform a nonlinear optimization to determine the $\phi(\nu)$ that maximizes S_Γ. We compute the gradient of the sharpness with respect to the phase values as follows:

$$\frac{\partial S_\Gamma}{\partial \phi(\nu)} = \sum_{x,y} w(x)\frac{\partial \Gamma[I(x, y)]}{\partial I(x, y)}\frac{\partial I(x, y)}{\partial \phi(\nu)}, \quad (8)$$

where[4]

$$\frac{\partial I(x, y)}{\partial \phi(\nu)} = \frac{\partial |g(x, y)|^2}{\partial \phi(\nu)}$$

$$= (2/N)\text{Im}[g^*(x, y)G(x, \nu)\exp(i2\pi\nu y/N)] \quad (9)$$

and $g^*(x, y)$ is the complex conjugate of $g(x, y)$. Performing the y summation yields

$$\frac{\partial S_\Gamma}{\partial \phi(\nu)} = (2/N)\sum_x w(x)$$

$$\times \text{Im}\left(G(x, \nu)\left\{\text{FT}\left[g(x, y)\frac{\partial \Gamma[I(x, y)]}{\partial I(x, y)}\right]\right\}^*\right). \quad (10)$$

Examples of the partial derivative of the point nonlinearity are

$$\frac{\partial \Gamma[I(x, y)]}{\partial I(x, y)} = \beta[I(x, y)]^{\beta-1} \quad (11)$$

for the power law of Eq. (7) and

$$\frac{\partial \Gamma[I(x, y)]}{\partial I(x, y)} = \ln[I(x, y)] + 1 \quad (12)$$

for $\Gamma[I(x, y)] = I(x, y)\ln[I(x, y)]$, negative (Shannon) entropy. Note that maximizing $I(x, y)\ln[I(x, y)]$, the negative of entropy, is equivalent to minimizing entropy.[6] Reference 4 considers the case of the most often used sharpness metric, the power law with $\beta = 2$. Other papers using sharpness metrics did not include the efficient calculation of the gradient, thereby limiting their usefulness to very-low-order phase errors (e.g., quadratic or cubic). The efficient calculation of the gradient requires just two fast Fourier transforms for all N of the partial derivatives.

The expressions for the sharpness metric and its gradient can be used in a nonlinear optimization algorithm, such as conjugate gradient, to find the phase-error estimate that maximizes the sharpness.

The description above treats each sample of the phase, $\phi(\nu)$, as an independent parameter, which can be referred to as a nonparametric phase estimate. As an alternative, one can express the phase error in terms of a basis-set expansion,

$$\phi(\nu) = \sum_{j=1}^J a_j L_j(\nu), \quad (13)$$

where, for example, the $L_j(\nu)$ could be the first J Legendre polynomials, and optimize over the coefficients a_j, for a parametric phase estimation. Then the gradient of the sharpness metric with respect to the coefficients is given by the chain rule for partial derivatives,

$$\frac{\partial S_\Gamma}{\partial a_j} = \sum_{x,y} w(x)\frac{\partial \Gamma[I(x, t)]}{\partial I(x, y)}\sum_\nu \frac{\partial I(x, y)}{\partial \phi(\nu)}\frac{\partial \phi(\nu)}{\partial a_j}$$

$$= (2/N)\sum_\nu L_j(\nu)\sum_x w(x)$$

$$\times \text{Im}\left[G(x, \nu)\left(\text{FT}\left\{g(x, y)\frac{\partial \Gamma[I(x, y)]}{\partial I(x, y)}\right\}\right)^*\right]$$

$$= (2/N)\sum_\nu L_j(\nu)\frac{\partial S_\Gamma}{\partial \phi(\nu)}, \quad (14)$$

which is just the projection of the nonparametric gradient onto the basis set.

In a typical gradient search technique, one computes the gradient, then computes from that a search direction, and then performs a line search to find the minimum of the objective function along the search direction. These three steps are repeated a few times until no further progress is made. The dominant burden of computing the entire gradient is two 1-D fast Fourier transforms for each range bin: one to compute $g(x, y)$ and the other ex-

plicit in Eq. (10). The gradient of Eq. (14) requires the J additional projections onto the basis set (N multiplies and adds). The dominant burden of the line search is the computation of $g(x, y)$ in Eq. (6), which again requires a 1-D fast Fourier transform. Typically three evaluations along the line search are adequate.

3. CUTS THROUGH METRIC SPACE

Given that we have an efficient gradient search algorithm for optimizing a sharpness metric for any point nonlinearity, the question is what is the best metric (point nonlinearity) to use. We answered this question by experimenting with many metrics on many types of images. We found that which was the best metric depended on the characteristics of the scene.

Simulations were performed starting from original images without phase errors and then adding known phase errors. The original images for an initial set of experiments are shown in Figs. 1 and 2. The random-clutter backgrounds consist of uniformly distributed Gaussian random numbers. A Taylor weighting function filtered the images in the Fourier domain. Various phase errors $\alpha \phi_e(\nu)$ were added to the phase of the signal history to blur the imagery. The phase errors were normalized such that the factor α is in units of radians rms (root mean squared). The types of phase errors used were quadratic, sixth-order (shown in Fig. 3), and random.

In a first set of experiments, we added a phase error $\alpha \phi_e(\nu)$ to the signal history and computed the sharpness of the resultant blurred image. The base phase error $\phi_e(\nu)$ was normalized to have unity standard deviation. To give an idea of the effect of α on blurring, a quadratic phase error having standard deviation of n radians rms causes a smearing of the image by $6\sqrt{5}n/\pi \approx 4.27n$ resolution elements. Good imagery requires satisfying the Rayleigh 1/4-wave criterion, equivalent to 0.468 rad rms, or the Marechal 1/14-wave rms criterion, equivalent to 0.449 rad rms. For imagery of superb quality, one might require 1/10 those amounts of phase error. Figures 4–6 show example plots of the sharpness metrics for the sixth-order phase error as a function of α for three different types of scenes and for different metrics. These plots show cuts through the multidimensional metric space, which is a function of the vector $\phi(\nu)$, in the direction of the particular sixth-order phase error $\phi_e(\nu)$. They give us a feeling for the structure of the space and allow us to see whether local extrema might be present. If local extrema exist, then the nonlinear optimization algorithm could get trapped in one of them and stagnate without finding the true solution. The top rows of plots show the metrics over a very wide range of α, and each bottom row is a version of the top row, zoomed in on the peaks of metrics, near $\alpha = 0$. The value of α at which a curve peaks would indicate the residual rms error that would be present after optimization is complete.

Note that a local maximum in one of these cuts through the metric space does not necessarily correspond to a local maximum in the multidimensional space; a path to the global maximum might exist by other paths. Furthermore, a smooth curve leading monotonically to the global maximum along one of these cuts does not necessarily
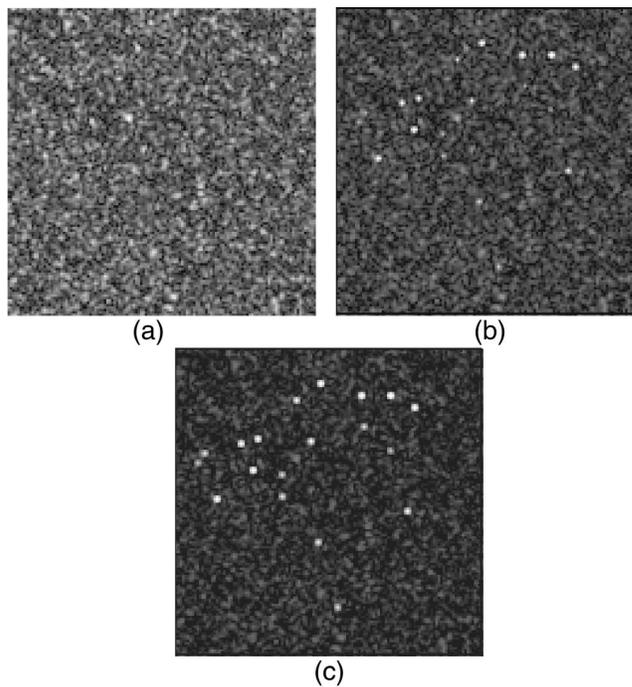
Fig. 1. Simulated images of point targets plus random-clutter background. Ratio of energy of point targets to energy of background: (a) 0, (b) 0.1, (c) 1.
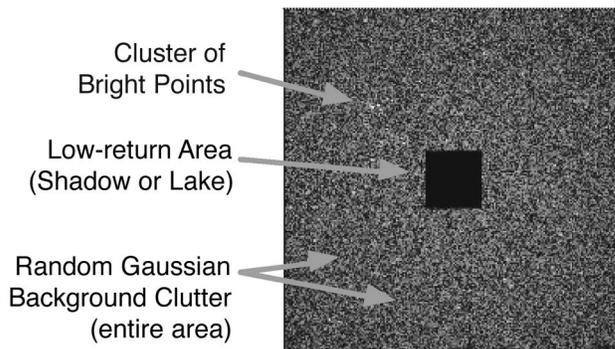
Fig. 2. Simulated point targets plus random-clutter background plus no-return area.
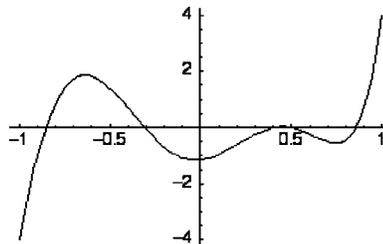
Fig. 3. Sixth-order phase error used for simulation experiments.

mean that an algorithm, even if starting on one of the points on the curve, cannot get trapped in a local minimum. The gradient in the multidimensional space will almost surely stray off that cut and could lead to a local maximum elsewhere. Despite the fact that these cuts through the metric space prove nothing absolutely, they

are nevertheless useful indicators of whether local maxima or difficulties in convergence are likely to occur.

The three metrics for these plots were power laws of the form of Eq. (7), with values of β = 0.5, 2.0, and 5.0. Since for β = 0.5 the true solution (α = 0) is at its minimum, we would minimize this metric or maximize its negative.

Figure 4 shows the case of bright point targets plus random clutter, shown in Fig. 1, with target/clutter energy ratio of 1.0. This image is very easy to focus for any algorithm, since bright, isolated points, often referred to as prominent points, dominate it. All of the metrics have single local extrema near the true solution, indicating that a nonlinear optimization algorithm is likely to have an easy time converging to a good estimate. The minimum for β = 2 was exactly at the true solution, and the metric is a very smooth function, suggesting that the searching routine would find the solution with a minimum of computation. In contrast to this, for β = 0.5, the

local global extremum is at α = 0.055 (rad rms), and the metric is not so smooth for larger values of α, suggesting that the searching routine would take longer to find the solution. For this easy object, we would expect all sharpness metrics to work reasonably well but higher powers β to work better.

Figure 5 shows the same thing for the case of target/ background ratio = 0.01, for which the image is dominated by the uniform clutter, but the bright points still play a role. In this case the low power β = 0.5 has many local extrema along the cut, and the peak value is far from the true solution, so it would perform very poorly. For β = 2 and 5, the peak was off by only α = 0.01 and 0.03, respectively, but for α > 5, there are several local-extrema. For this object, the larger powers β are necessary, but they could easily have problems with local maxima.

Figure 6 shows a case for an image in which there are no bright points but there is a large no-return area in the
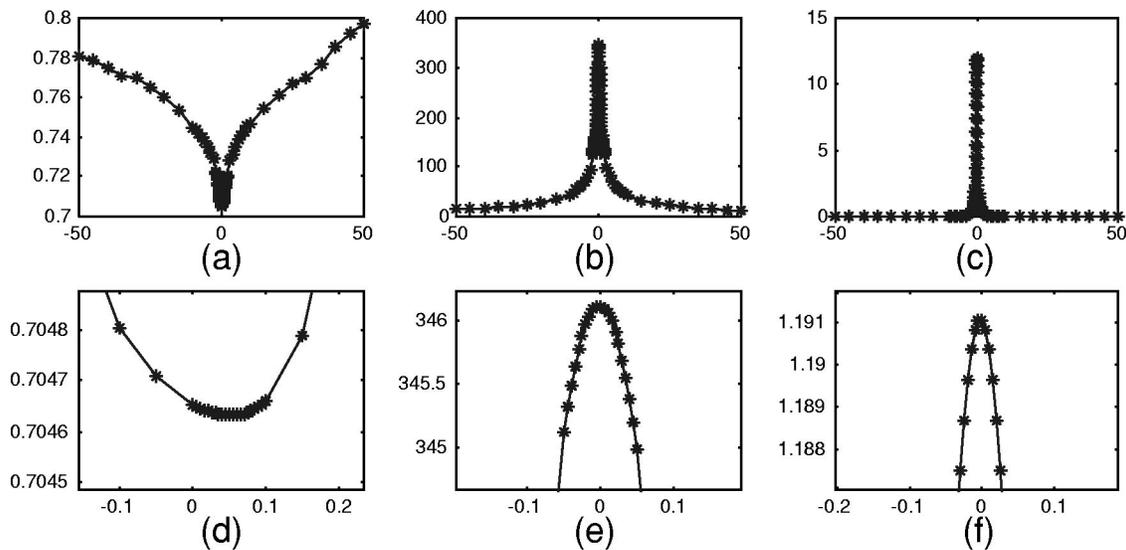
Fig. 4.   Plots of sharpness metrics versus rms phase error α (in radians) for a few bright points plus clutter background, with target/ background ratio of 1.0. All are for a sixth-order polynomial phase error and for point nonlinearity $\Gamma[I] = I^{\beta}$. (a) β = 0.5, (b) β = 2.0, (c) β = 5, (d)–(f) expanded versions of (a)–(c), respectively.
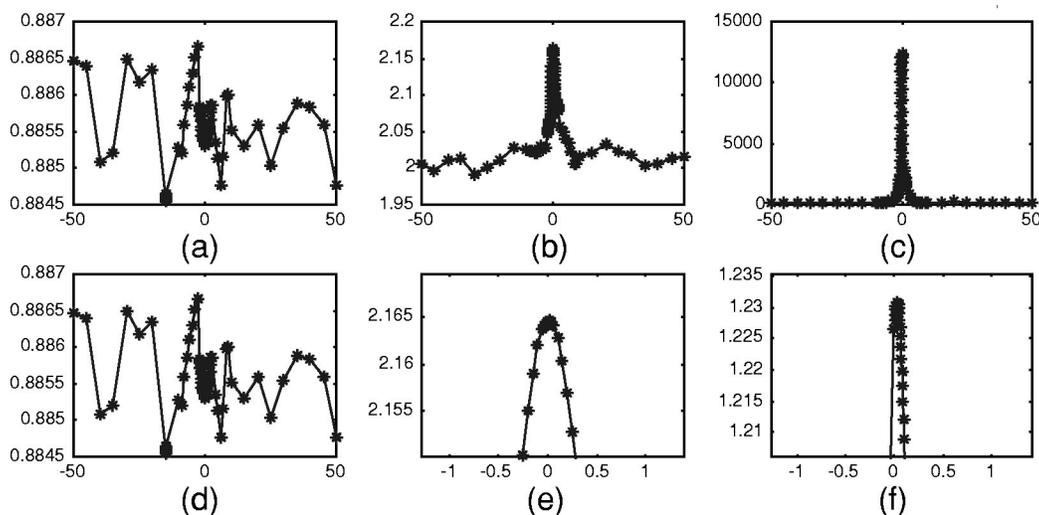
Fig. 5.   Same as Fig. 4, but with target/background ratio of 0.01.
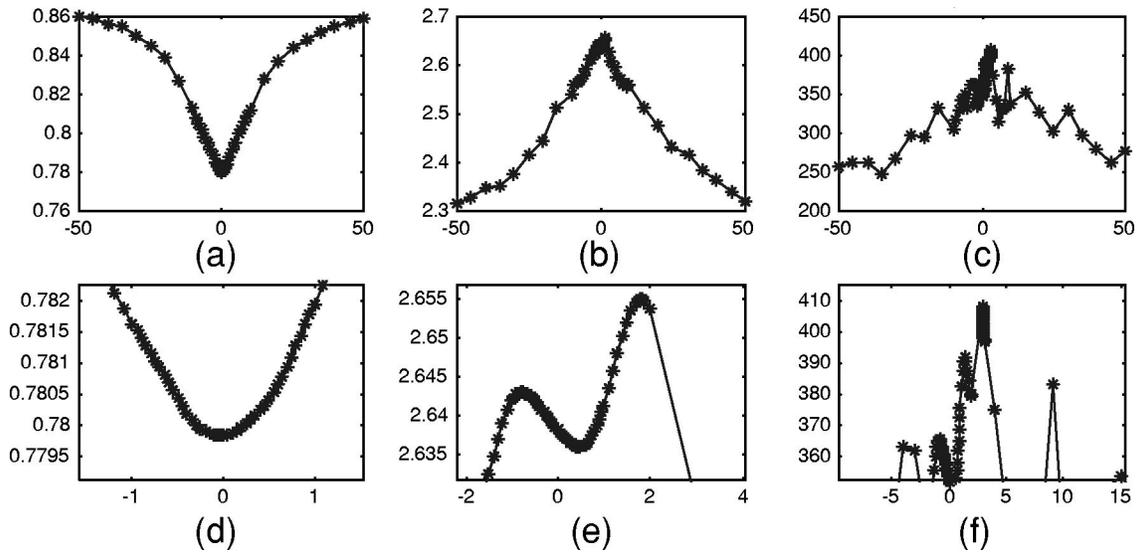
Fig. 6.   Same as Fig. 4, but with no bright points and with a 128×128 no-return area embedded in a 384×384 clutter background.

image, as shown in Fig. 2.   In this case $\beta = 5$ works very poorly, and $\beta = 2$, while better, has the peak at $\alpha = 1.8$ and has significant local maxima.   For this object, $\beta = 0.5$ works the best, having no local minima, and has its peak at $\alpha = -0.075$.

From these examples we see that which metric performs best depends highly on the content of the scene, with large powers of β best for images containing prominent points and small powers of β best for images containing dark regions with no prominent points.   In Section 4 we explain why the metrics act in this way.

4.   IMPORTANCE OF THE SECOND DERIVATIVE

The point nonlinearity $\Gamma[I]$ determines the nature of the objective function.   A key to understanding its effect is illustrated in Fig. 7, which shows a plot of a point nonlinearity $\Gamma[I]$ having a positive second derivative

$$\Gamma''[I] = \frac{\partial^2 \Gamma}{\partial I^2}. \tag{15}$$

As the phase-error estimate is changed, the sum of the intensities remains constant, according to conservation of energy.   Suppose that we start with an image having two pixels with the same intensity, I_0, making their average intensity I_0.   If the phase correction causes the energy ΔI from one of these two pixels to be transferred to the other, the new pixel intensities become $I_1 = I_0 - \Delta I$ and $I_2 = I_0 + \Delta I$, whose average value remains I_0.   However, if $\Gamma''[I] > 0$ in the interval $[I_1, I_2]$, then one can see from Fig. 7 that $\Gamma[I_1] + \Gamma[I_2] > 2\Gamma[I_0]$.   That is, the sharpness, Eq. (6), increases when the intensity values are increasingly different from one another.   More generally, suppose that $\Gamma''[I] > 0$ in the interval $[I_1, I_2]$, making $\Gamma[I]$ a convex function in that region, and the values $I(x, y) \in [I_1, I_2]$ deviate from their mean, $\bar{I}$.   Then (ignoring the weighting function) by Jensen's inequality[7,8] we have

$$S_\Gamma = \sum_{x,y} \Gamma[I(x, y)] > MN\Gamma[\bar{I}]. \tag{16}$$

That is, the sharpness increases when the values of the intensity deviate from their mean.   This means that choosing a phase correction that increases sharpness will stretch the histogram of the intensity values and increase their contrast and standard deviation.   The larger $\Gamma''[I]$ is, the greater will be this effect.

As illustrated in Fig. 8, for $\Gamma[I] = I^2$, the classic squared-intensity sharpness, $\Gamma''[I] = 2$, a constant.   This nonlinearity gives the same weight to stretching the histogram for low values of I as it does for high values of I.   More generally, for $\Gamma[I] = I^\beta$, $\Gamma''[I] = \beta(\beta - 1)I^{\beta-2}$.   For powers $\beta > 2$, $\Gamma''[I]$ increases with increasing I.   For $\beta \gg 2$, $\Gamma''[I]$ is much larger for larger I.   This means that the sharpness metric will give much more weight to stretching the histogram for large values of I than for small values of I.   That is, it will give much more weight to making the bright pixels brighter than to making the dark pixels darker.   At an extreme, as $\beta \to \infty$, maximizing the sharpness metric will concentrate only on making brighter the single brightest pixel (maximizing the maximum value).   Conversely, for powers $\beta < 2$, $|\Gamma''[I]| = |\beta(\beta - 1)|/I^{2-\beta}$ decreases with increasing I.   In this case, the sharpness metric will give more weight to stretching the histogram for smaller values of I.   That is, it will give more weight to making the dark pixels darker than making the bright pixels brighter.   For $\beta = 0.5$, as used in the initial experiments described earlier, $\Gamma''[I] = -2.5/I^{1.5}$, which would emphasize darkening shadows.   We would maximize the negative of this sharpness metric since this second derivative has a negative sign.   For $\beta = 1 + \epsilon$, where $0 < \epsilon < 1$, the second derivative $\Gamma''[I] = (1 + \epsilon)\epsilon/I^{1-\epsilon}$ is positive and decreases with increasing intensity.   Figure 8 shows four different point nonlinearities and their second derivatives, illustrating how the values of $\Gamma''[I]$ differ with I for different point nonlinearities.

Included in Fig. 8 is the case of negative Shannon entropy, $\Gamma[I] = I \ln(I)$.   Its second derivative, $\Gamma''[I] = 1/I$,
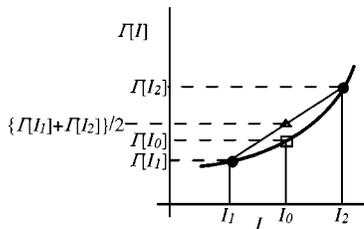
Fig. 7. $\Gamma[I]$, the point nonlinearity of the intensity. If $\Gamma''[I]$ is positive, then the sharpness, the sum over $\Gamma[I]$, will increase as the values of I are spread out.
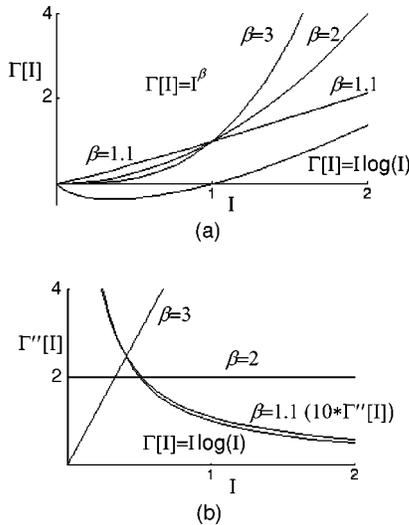
Fig. 8. (a) Four point nonlinearities and (b) their second derivatives.

is approximately equal to a constant times the second derivative of $\Gamma[I] = I^{1+\epsilon}$ (given above) when $0 < \epsilon \ll 1$, even though the plots of these two point nonlinearities themselves appear to be very different. Hence, if the theory that the second derivatives drive the behavior of sharpness maximization is correct, then the behavior of these two point nonlinearities for sharpness maximization should be nearly the same. As we will see later, this is indeed the case. Later we will also refer to exponential entropy, by which we mean $\Gamma[I] = -I \exp(1-I)$,[9] for which $\Gamma''[I] = -(I-2)\exp(1-I)$. When I is normalized to have values less than unity, $\Gamma''[I]$ decreases with increasing I, which, like maximizing negative Shannon entropy, emphasizes darkening of shadows.

Stretching the histogram of the image is valid for phase-error correction since the blurring that occurs because of a phase error smears out bright pixels, reducing their intensity toward the mean, and fills in dark areas, increasing their intensity toward the mean. SAR images often contain extremely bright points if corner reflectors are present in the scene. They also usually contain low- (or no-) return areas. These include smooth objects, such as surfaces of water or paved roads, which reflect the energy away from the SAR, and shadow areas. Since the viewpoint of a range-angle SAR image is perpendicular to the line of sight, image pixels behind tall three-dimensional objects that rise above the ground plane will be shadowed by the tall objects and will ideally have zero

intensity. For inverse SAR imaging of moving targets, often the background is inherently dark. Focusing an image by any means will make the bright points brighter and the dark areas darker (which is usually desirable); and the point nonlinearity used in the image sharpening algorithm will determine the extent to which either of these two things is emphasized. When the image contains very bright pointlike objects, then the analysis above suggests that point nonlinearities that have $\Gamma''[I]$ increasing with increasing I (e.g., with higher powers β) would perform better. However, when the image contains no bright points (such as a forest or fields) and one must rely on the shadows to create image contrast, then the analysis suggests that point nonlinearities that have $\Gamma''[I]$ decreasing with increasing I (e.g., with powers β near or below unity) would perform better.

5. DESIGNER METRICS

We have found that the second derivative of the point nonlinearity defining a sharpness metric determines its action on the reconstructed image. Hence one can design a point nonlinearity to take advantage of the properties of a specific image type. Examples are as follows. Suppose that one expects bright points to be in the image, making a power law, Eq. (7), with $\beta = 4$ advantageous. However, as can be seen from Fig. 8(b), with these larger values of β, $\Gamma''[I]$ approaches zero as I approaches zero, causing the metric to pay little attention to making the dark pixels darker. A seeming improvement would be to add an offset to the second derivative of the point nonlinearity,

$$\Gamma''_{D1}[I] = (I - \gamma I_0)^2, \qquad (17)$$

which is shown in Fig. 9(a). This curve rises as it approaches zero, giving more emphasis to making shadows darker. The corresponding point nonlinearity, gotten by integrating twice, is

$$\Gamma_{D1}[I] = (1/12)(I - \gamma I_0)^4, \qquad (18)$$

which is shown in Fig. 9(b). Suppose that one wants to give even more emphasis to making shadows darker. One could increase the second derivative near $I = 0$ with the following:

$$\Gamma''_{D2}[I] = \frac{(I - \gamma I_0)^2}{I + \epsilon}, \qquad (19)$$

shown in Fig. 9(c). A small constant ϵ is included in the denominator to avoid having $\Gamma''_{D2}[I]$ go to infinity at $I = 0$. $\Gamma_{D2}[I]$, shown in Fig. 9(d), is a complicated expression involving terms in I, I^2, I^3, $I \ln(I + \epsilon)$, and $\ln(I + \epsilon)$. To modify this metric to further emphasize bright points, one could use a larger power in the numerator,

$$\Gamma''_{D3}[I] = \frac{(I - \gamma I_0)^4}{I + \epsilon}, \qquad (20)$$

which is shown in Fig. 9(e). $\Gamma_{D3}[I]$, shown in Fig. 9(f), is a complicated expression involving terms in I, I^2, I^3, I^4, I^5, $I \ln(I + \epsilon)$, and $\ln(I + \epsilon)$.
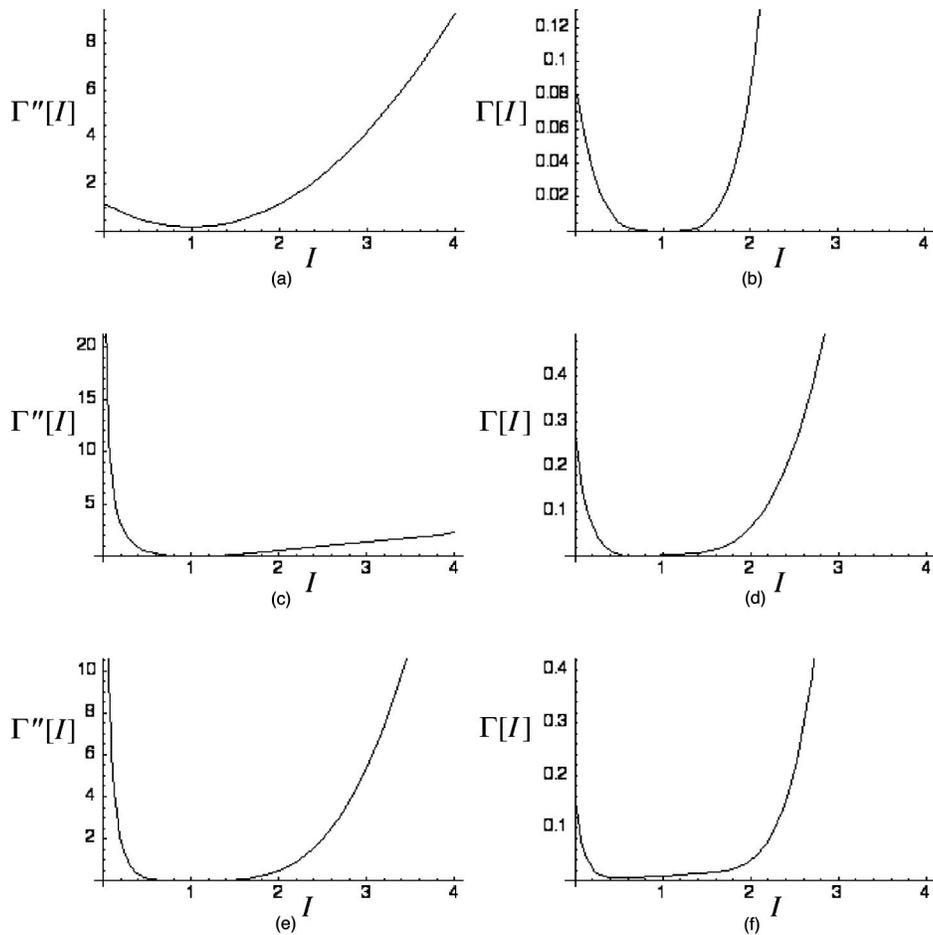
Fig. 9.   Examples of designer metrics and their second derivatives.   (a) Γ''_{D1}, (b) Γ_{D1}, (c) Γ''_{D2}, (d) Γ_{D2}, (e) Γ''_{D3}, (f) Γ_{D3} (defined in the text).

6. CONVERGENCE OF DIFFERENT METRICS

Next we show the results of computer simulation experiments employing the image-sharpening metrics discussed above.   Figure 10 shows real SAR images used as the objects for these experiments.   Again, a variety of phase errors were added to their Fourier transforms, and the performance of different metrics was compared.

Figure 11 shows an example of focusing a relatively difficult object, Trees.   We used the nonparametric version of the algorithm, not taking advantage of the fact that the phase error was smooth.   So even though the phase error was sixth order, the algorithm would allow for correcting a very-high-order phase error.   The same thing was done for the other examples shown later.   Nevertheless, the reconstructed image is excellent.   A popular algorithm based on prominent points[10] failed for this case.

For this reconstruction example, Fig. 12 shows plots of scaled metrics (objective functions) versus phase error for the following algorithms: power-law sharpness with β = 1.1 and β = 5, and metric D1 with γ = 1 and γ = 100.   Because the different metrics have values that differ by orders of magnitude, they cannot be conveniently plotted together, so we plot here scaled metrics,

$$E_S = \frac{S - S_o}{S_i - S_o},\tag{21}$$

where S is the normal sharpness metric, S_o is the value of S for the original image with no phase errors, and S_i is the value of S for the initial smeared image.   This scaled metric starts at unity where $S = S_i$ and then, as the iterations progress, ideally decreases toward zero as S approaches the ideal value S_o.   The scaled metric decreases monotonically with iteration number.   The appearance of plateau regions followed by sudden decreases occurs because of the nature of conjugate-gradient search.   The conjugate gradient implicitly accumulates information about the second partial derivatives of the sharpness metric with respect to the search parameters as the iterations progress.   However, at the same time the solution is moving into a different area of parameter space where the curvature of the metric is different, making the second-derivative information obsolete.   Consequently, it is necessary to restart the conjugate gradient (with one iteration of steepest descent) every once in a while.   At these times the sharpness metric improves rapidly for several iterations.   For these examples we restarted the conjugate-gradient search every 50 iterations.   We define an iteration as one evaluation of the metric.   Typically three evaluations of the metric are computed for each line search direction after the gradient and search direction are recalculated.   Occasionally the line search that we employed, which is from the *fminu* optimizer in

the Optimization Toolbox of Matlab™, takes a significantly larger number of metric evaluations.

For some of the metrics investigated, the powers were large enough to cause numerical problems in the optimization routines. This problem was initially addressed by scaling the objective function and the gradients by a constant, to give them values closer to unity. This yielded much better convergence properties and alleviated some of the numerical issues. A better fix to the problem was to scale the complex image such that the intensity values, when raised to the powers in the metric, would be in an acceptable range. This prevented the sums of very large or very small numbers from occurring. An empirical approach was taken to determine the scaling of the image intensity that yielded metric values in the desired range. The typical range was [1, 100]. This approach provided a more acceptable result and was used throughout most of the analysis.

For this example, both metrics D1 with $\gamma = 1$ and the power law with $\beta = 5$ converged to scaled objective functions having negative values. This means that the sharpness of the corrected image estimate is significantly greater than the sharpness of the original (ideal) image. The sharpness metric is the measure by which the algorithm tries to focus the image. However, it is possible for a metric to improve even though the image is becoming increasingly blurred. One example of how this can happen is the case when the image is oversharpened. Suppose that there is no weighting on the range bins and that one range bin, x_0, contains a few relatively bright scatterers. The signal history for the blurred image in that range bin is

$$G_d(x_0, \nu) = F(x_0, \nu)\exp[i\phi_e(\nu)]$$
$$= |F(x_0, \nu)|\exp[i\phi_e(\nu) + i\psi(x_0, \nu)], \quad (22)$$
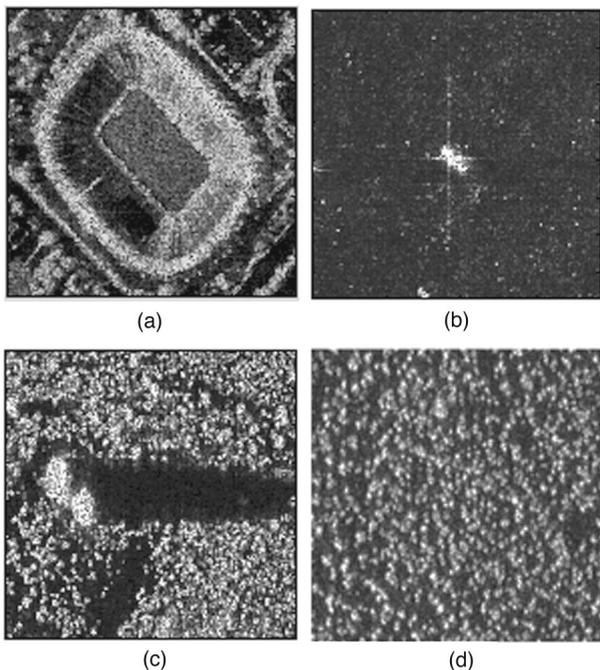
Fig. 10.   Real SAR images used in simulations, taken with Veridian/ERIM's DCS: (a) Stadium, (b) Target, (c) Shadow, (d) Trees.

(a)   (b)
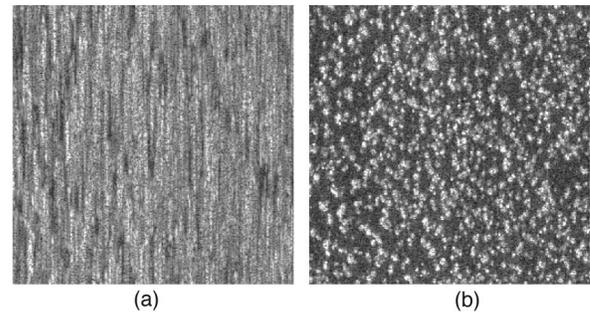
Fig. 11.   Image-focusing example.   (a) Image of Trees smeared by 20-rad rms of sixth-order phase error, (b) image focused after 140 iterations of image sharpening with $\beta = 2$ power-law metric.
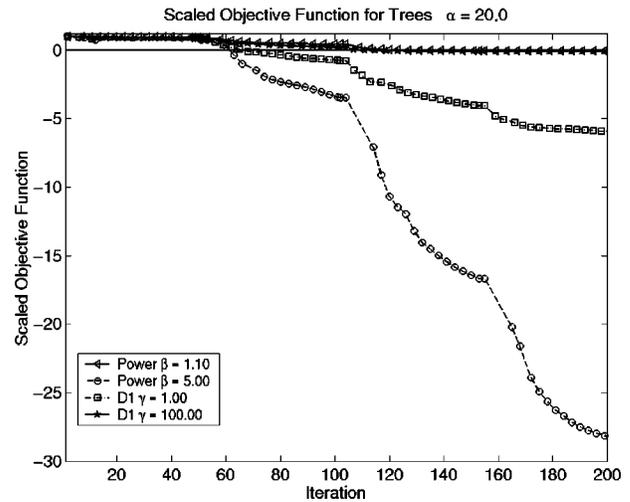
Fig. 12.   Scaled objective function versus iteration number for Trees.

which has phase $\phi_e(\nu) + \psi(x_0, \nu)$.   A phase-error estimate of

$$\phi(\nu) = -[\phi_e(\nu) + \psi(x_0, \nu)] \quad (23)$$

will cause the phase of $G(x_0, \nu)$ to be zero, making $G(x_0, \nu)$ real and nonnegative. The corresponding image will then be dominated by a strong delta-function-like impulse response in the center of the range bin. That is, the energy of that range bin, which should be distributed among multiple scatterers, is mostly concentrated into a single scatterer. This improper phase-error estimate will blur the other range bins, decreasing their sharpness, but will greatly increase the sharpness of range bin x_0. If that range bin is bright enough, then it is possible that the algorithm will converge to this false local extremum. There are a variety of ways to reduce the probability of getting trapped in such potential false solutions. For example, one can normalize the energy of the range bins, using the weighting function $w(x)$, so than none of them can dominate over all the others. Another way would be to restrict the phase estimate to being a low-order polynomial by using the parametric form of the gradient search as in Eq. (14), at least for early iterations. For the example shown in Fig. 12 no such measures were taken, and the two metrics for which oversharpening happened

place so much emphasis on making the brightest point brighter that they converge to an oversharpened image.

The possibility of oversharpening also means that, when comparing sharpness metrics, one should not base success only on the degree to which a given sharpness metric improves. Another metric, based on the truth, should be used. One way to do this, when the phase errors are digitally simulated, is to compute the mean squared difference between the true phase error and the estimated phase error. The mean squared error is not the most reliable measure of quality as perceived by a human; but, for a given phase-error type, it is highly correlated with perceived image quality and is often used in evaluating algorithm performance. This use is problematical, however, when the phase error is random or the nonparametric phase estimation is employed, because the phase is wrapped (computed modulo 2π). There are also problems with the fact that the metrics are insensitive both to constant and to linear phase terms. Linear phase terms just shift the image without smearing them. To avoid these difficulties, we employ the invariant error metric.[11] Since the Fourier magnitude is known, the appropriate invariant metric is that given in Eq. (12) of Ref. 11:

$$E^2 = \min_{a,x_0,y_0} \left\{ \frac{\sum |e^{ia}g(x-x_0, y-y_0) - f(x,y)|^2}{\sum |f(x,y)|^2} \right\}$$

$$= \frac{r_{gg}(0,0) + r_{ff}(0,0) - 2\max_{x_0,y_0}|r_{fg}(x_0,y_0)|}{r_{ff}(0,0)}, \quad (24)$$

where r_{fg} is the cross correlation of $f(x,y)$, the ideal image (with no added phase errors), with $g(x,y)$, the image estimated after correction of the phase errors. It is a normalized mean squared difference between the reconstructed image and the ideal image, allowing for both constant phase differences and linear phase differences (image translations). We report the square root of this metric, making it a normalized rms error.

Figure 13 shows the normalized rms errors corresponding to the results given in Fig. 12. From this we see that the two metrics which appeared to improve the most in
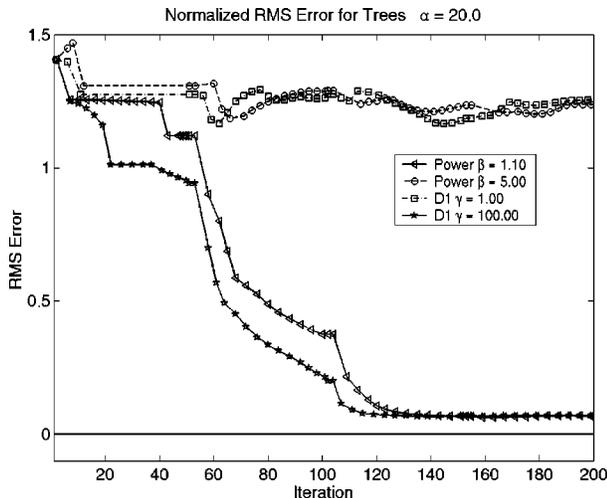
Fig. 13. Normalized rms error versus iteration number for Trees.
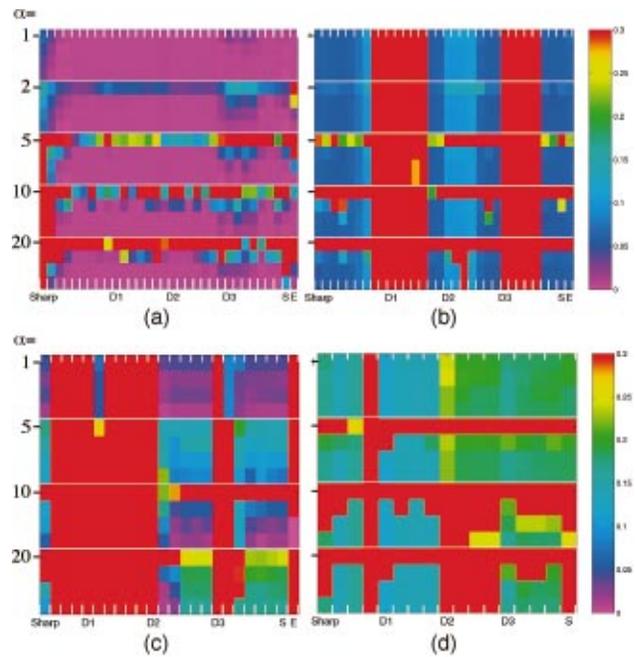
Fig. 14. Color plot showing performance of various metrics (see text). (a) Target, with quadratic phase error; (b) Trees, with quadratic phase error; (c) Clutter Plus Shadow, with sixth-order phase error; (d) Clutter Plus Points, with target/clutter ratio 0.1 and quadratic phase error. Metrics for (a) and (b) (left to right): Power law $\beta = 0.45, 0.6, 0.95, 1.1, 2, 2.5, 3, 4, 5$; D1, D2, and D3, $\gamma = 0.001, 0.01, 0.1, 1, 10, 100, 1000$; Shannon entropy, exponential entropy. Metrics for (c) (left to right): Power law $\beta = 1.1, 2, 2.5, 4$; D1, D2, and D3, $\gamma = 1, 10, 100, 1000$; 2500, 5000; Shannon entropy, exponential entropy. Metrics for (d) (left to right): Power law $\beta = 1.1, 2, 2.5, 4$; D1, D2 and D3, $\gamma = 100, 1000$; 2500, 5000; Shannon entropy.

terms of their objective functions (D1 with $\gamma = 1$ and power law with $\beta = 5$) in fact were performing very poorly, their rms errors decreasing only slightly. The two other metrics shown, D1 with $\gamma = 100$ and power law with $\beta = 1.1$, both converged to a much lower error and performed relatively well.

We performed similar reconstruction experiments with several metrics and parameters for each metric, several phase-error types, and sizes of phase errors and for a few types of SAR images (cultural versus rural), totaling thousands of cases. To evaluate the relative performance of the different metrics for so many cases, comparing plots of rms error versus iteration number is inefficient. Also, because searches through high-dimensional spaces can at times be chaotic (i.e., small changes in the initial conditions can result in large changes in the final result), it is necessary to perform numerous experiments to obtain statistically meaningful results. To compare so many cases, we used color plots, examples of which are shown in Fig. 14. Each colored block in a color plot represents the residual rms error of the corrected image at a given stage in the reconstruction process, the corresponding value of which is given in the color bar to the right. For residual errors much less than unity, the normalized rms error reported here is approximately equal to the residual rms phase error in radians.[11] Red represents a residual error of 0.3 or greater. Recall that a residual error of 0.45 is needed to match the Marechal or the Rayleigh criteria, so

even this (red) level is not necessarily a bad residualphase error. This demanding scale was used so that it could better distinguish between the merely good and the excellent results. For imagery of excellent quality, one would want to achieve, say, 0.05, represented by deep blue. Light violet represents negligibly small residual error. Each column represents four or five different reconstruction experiments with initial phase errors of different sizes, indicated by α (in radians rms), shown to the left. In each case we ran four groups of thirty iterations each of conjugate gradient, where we restarted with one iteration of steepest descent at the beginning of each group of iterations. The first four blocks of color in each column represent the residual error after each of the four blocks of iterations for a starting phase error with α = 1. The next four blocks of color represent the residual error for each of four blocks of iterations for the next indicated starting value of α, and so on. Each column is for a different metric or a metric with a different parameter. For example, in Fig. 14(a), the first nine columns are for the power law with nine different values of the power β, the next seven columns are for designer metric D1 with seven different values of the parameter γ, and so on as indicated in the figure caption. In total, Fig. 14(a) shows the results of 160 focusing experiments (5 values of initial α and 32 different metrics), each of which is shown at four stages. Similar series of experiments were performed for several different image types and for three different phase-error types (quadratic, sixth-order, and random). Metrics can be readily compared by seeing which column reaches the lowest value of residual error in the shortest time.

Figure 14(a) shows that for the case of Target, which is easy to focus, almost all the metrics arrived at excellent results in relatively few iterations. For the power-law metric, powers of β near 2 performed significantly better than smaller powers. The entropies converged more slowly than the other metrics.

Figure 14(b) shows, for the more difficult case of Trees, that none of the metrics yielded extremely low residual errors. We believe that this happened because the image with which we started had a small residual quadratic phase error. For the metrics that worked well, the algorithm converged to a better image than this original image, but our residual-error calculation counted this as a nonzero residual error. From these results we see that for power-law sharpening, lower powers worked best and the powers β = 4 and above worked poorly. Other metrics that worked well were both entropies and all three designer metrics with large values of γ.

Figure 14(c) shows that for the most difficult case of Clutter Plus Shadow (as in Fig. 2 but without the bright points), the results were very mixed. Which metric worked best depended on the size of the initial phase error, which is due to the sometimes-chaotic nature of nonlinear optimization searches. The metric that worked best seemed to be D2, which was designed specifically to work well with an image containing shadows. Almost as good were the power law with a small value of β = 1.1 and Shannon entropy. These results are consistent with the theory of the second derivative of the point nonlinearity. The metrics that worked best have relatively large second derivatives for small values of the intensity, which makes them concentrate on making the shadows darker, and shadows are an important feature of this image.

Figure 14(d) shows that, for the case of Clutter Plus Points (as in Fig. 1, with a target/clutter ratio of 0.1), the best metrics were power law with larger values of β (=2.5 or 4), D1, and D3. Again these results are consistent with the theory of the second derivative. The metrics that worked best have relatively large second derivatives for large values of the intensity, which makes them concentrate on making the bright points brighter, and bright points are an important feature of this image. For the case of the same object, but with larger target/clutter ratio, the same trends were observed, but all the metrics converged to much lower residual errors. For a target/clutter ratio above 0.3, all metrics resulted in excellent reconstructions.

Trends similar to those shown in Fig. 14 held for all three types of phase errors employed: quadratic, sixth-order, and random. The metrics that worked best for the image of Stadium were the same as those that worked best for the image of the Target, which is to be expected because both contain bright pointlike targets. Similarly, the metrics that worked best for the image of the Trees were the same as those that worked best for the image with Shadow. Again, the best metric depended not much on the specific details of the particular image but on the general character or statistics of the image.

7. SIMILARITIES OF CERTAIN METRICS

Examination of Fig. 14 shows that certain pairs of metrics tend to perform very similarly to one another. Previously we mentioned that Shannon entropy is equivalent to a power law with β near unity. In addition, expanding Eqs. (17) and (18), we see that metric D1 approaches the power law with β = 2 when $\gamma \ll 1$ and also approaches the power law with β = 2 for $\gamma \gg 1$. (Note that terms proportional to intensity I or that are constants do not change when the phase estimate is changed, so such terms should be ignored.) Expanding Eq. (19) and its integral taken twice, we find that metric D2 is similar to a power law with β = 3 for $\gamma \ll 1$ and to Shannon entropy, $I \ln I$, for $\gamma \gg 1$. Expanding Eq. (20) and its integral taken twice, we find that metric D3 is similar to a power law with β = 5 for $\gamma \ll 1$ and to negative Shannon entropy for $\gamma \gg 1$. These similarities are generally borne out by the results seen in the reconstruction experiments, examples of which are shown in Fig. 14. This experimental fact, that metrics having similar second derivatives also have similar performance in a focusing algorithm, confirms the validity of the argument that the second derivative is the key to understanding metrics.

8. PHASE-ERROR-INDEPENDENT SCENE STATISTICS

For correcting phase errors in coherent images by image sharpening, we have established that the best sharpness metric depends on the characteristics of the scene. An obvious approach would be to measure the statistics of the scene and apply a metric that is best for those scene statistics. However, in the presence of large phase er-
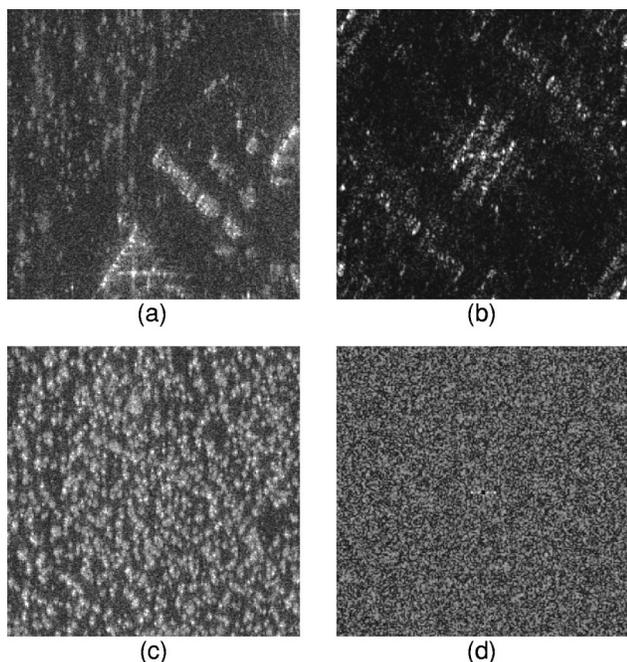
Fig. 15. Images and their autocorrelations. (a) SAR image of cultural objects, (b) autocorrelation of (a), (c) SAR image of Trees, (d) autocorrelation of (c).
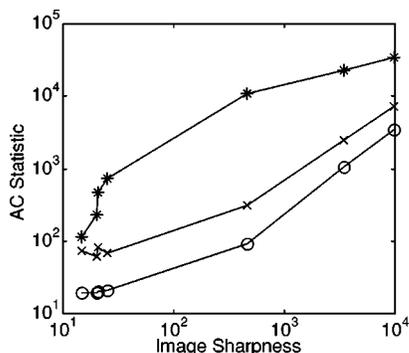
Fig. 16. Autocorrelation statistics versus image sharpness. * Kolmogorov–Smirnov; x, maximum value; ○, variance.

rors, it may be very difficult to determine the characteristics of the (unsmeared) scene from a severely smeared image of that scene. Our approach to this dilemma was to compute the statistics of scene-related quantities that are independent of the phase error. One of these quantities is the Fourier magnitude, which is unaffected by errors in the Fourier phase. Another of these quantities is the autocorrelation function, which is the inverse Fourier transform of the squared Fourier magnitude. As illustrated in Fig. 15, a highly structured scene has an autocorrelation that is highly structured, whereas a clutter-like scene has an autocorrelation that is highly random. We computed several statistics of the autocorrelation magnitude and the Fourier magnitude and compared them with the degree of structure of the focused scene. For the scene statistic we used the $\beta = 2$ power-law sharpness as the indication of the structure in the scene. This sharpness metric was normalized by dividing by the square of the energy in the scene. We computed the

maximum, the variance, and the Kolmogorov–Smirnov statistic (a measure of how much the statistics differ from Gaussian) of the Fourier magnitude and the autocorrelation magnitude away from the origin (where the autocorrelation has a delta function) and other statistics. As shown in Fig. 16, for the autocorrelation statistics we found that these quantities did correlate well with the actual sharpness of the (unsmeared) image. That, in turn, correlates well with the ease of performing autofocus. As mentioned earlier, these statistics can be used to determine which metric will perform better. In this case, performing better means converging faster. Another measure of performance is, given enough time to fully converge, the final residual error in the focused image. By this measure we found, through extensive experiments, that the lower powers and negative entropy tended to perform better than other metrics for most scene types.

9. CONCLUSIONS

In this paper we have explored the use of a variety of sharpness metrics for correcting phase errors in synthetic-aperture radar, a form of coherent imaging. It was found that different sharpness metrics worked better depending on the type of scene. Power-law metrics with larger powers tend to perform better with scenes having prominent scatterers, whereas power-law metrics with smaller powers tend to perform better with scenes having no prominent scatterers. A new theory for explaining these trends was developed: The behavior of a metric is determined by the second derivative of its point nonlinearity as a function of the image intensity. It was found both theoretically and experimentally that metrics having similar second derivatives gave similar results when used in an image-sharpening algorithm. For example, negative Shannon entropy acts very similarly to a power law with a power near unity. Metrics whose second derivative increases with increasing intensity emphasize making the bright points brighter, whereas metrics whose second derivative increases with decreasing intensity emphasize making the shadows and low-return areas darker. With these trends in mind, we can specify designer metrics that are tuned to the characteristics of a given image. For example, a metric designated D2 that was designed to work best on images containing shadow areas did in practice work particularly well on an image containing a large shadow area. Those image statistics can be approximately determined from the smeared image by examining either the autocorrelation of the image or the Fourier magnitude of the image, both of which depend on the structure of the ideal image but are independent of the phase error.

Portions of this work were presented in Ref. 12.

ACKNOWLEDGMENTS

The corresponding author can be contacted by email at fienup@optics.rochester.edu.

REFERENCES

1. R. A. Muller and A. Buffington, "Real-time correction of atmospherically degraded telescope images through image sharpening," J. Opt. Soc. Am. **64**, 1200–1210 (1974).
2. R. G. Paxman and J. C. Marron, "Aberration correction of speckled imagery with an image-sharpness criterion," in *Statistical Optics*, G. M. Morris, ed., Proc. SPIE **976**, 37–47 (1988).
3. F. Berizzi and G. Corsini, "Autofocusing of inverse synthetic aperture radar images using contrast optimization," IEEE Trans. Aerosp. Electron. Syst. **32**, 1185–91 (1996).
4. J. R. Fienup, "Synthetic-aperture radar autofocus by maximizing sharpness," Opt. Lett. **25**, 221–223 (2000).
5. J. L. Walker, "Range-Doppler imaging of rotating objects," IEEE Trans. Aerosp. Electron. Syst. **AES-16**, 23–52 (1980).
6. B. C. Flores, "A robust method for the motion compensation of ISAR Imagery," in *Intelligent Robots and Computer Vision X: Algorithms and Techniques*, D. P. Casasent, ed., Proc. SPIE **1607**, 512–517 (1991).
7. S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd ed. (Academic, New York, 1975), p. 249.
8. H. L. Royden, *Real Analysis*, 2nd ed. (MacMillan, New York, 1968), p. 110.
9. B. C. Flores, A. Martinez, and J. Hammer, "Optimization of high-resolution-radar motion compensation via entropy-like functions," in *Antennas and Propagation Society International Symposium Digest* (Institute of Electrical and Electronics Engineers, New York, 1993), pp. 1906–1909.
10. C. V. Jakowatz, Jr., and D. E. Wahl, "An eigenvector method for maximum likelihood estimation of phase errors in SAR imagery," J. Opt. Soc. Am. A **10**, 2539–2546 (1993).
11. J. R. Fienup, "Invariant error metrics for image reconstruction," Appl. Opt. **36**, 8352–8357 (1997).
12. J. R. Fienup and J. J. Miller, "Generalized image sharpness metrics for correcting phase errors," in *Signal Recovery and Synthesis*, Vol. 67 of Topics in Optics and Photonics Series (Optical Society of America, Washington, D.C., 2001), pp. 84–86.