MAXIMUM ENTROPY IMAGE RESTORATION REVISITED

Matthew Willis, Brian D. Jeffs and David G. Long

Department of Electrical and Computer Engineering
Brigham Young University, Provo, UT 84602
bjeffs@ee.byu.edu long@ee.byu.edu

ABSTRACT

This paper presents a new non-iterative, closed-form approximation to the maximum entropy (M.E.) image restoration method. A fast frequency domain implementation of this closed form approach is developed for the case of circular convolutional blur. This result dramatically reduces computational demands compared to conventional iterative M.E. algorithms such as MART. Some limitations and advantages of M.E. restoration are investigated, including its dismal performance for high resolution restoration of decimated or randomly sampled blurred observations.

1. INTRODUCTION

Maximum entropy (M.E.) image restoration has been widely used with significant success for many years. It remains a method of choice for radio astronomy image restoration applications. Though usually expressed as a deterministic constrained convex optimization problem, proponents of M.E. restoration have used information theoretic and quantum physics arguments to justify its use in a wide variety of applications [1][2].

We will use the following observation model,

$$y = Hx, \tag{1}$$

where $y : M \times 1$ and $x : N \times 1$ are vectors formed by column scanning the corresponding 2-D images (e.g. $x = vec\{x[m,n]\}$), and H is the doubly block Toeplitz convolution matrix corresponding to the 2-D point spread function, $h[m,n]$. It will be assumed that the rank of H is M, with $M \leq N$, so that given y, the solution for x is not unique unless some optimization criterion (such as maximum entropy) is employed. This rank deficiency also ensures that the strict equality constraint of equation (1) can be satisfied even in the presence of observation noise, though the solution will be perturbed. The corresponding M.E. optimization problem for image restoration is given by

$$\hat{x}_{ME} = \arg\max_{x \geq 0} -\sum_{i=1}^{M} x_i \ln x_i,$$
$$\text{such that } y = Hx. \tag{2}$$

This M.E. problem my be solved using general purpose non-linear constrained optimization computer algorithms, or one of several algorithms related to the multiplicative algebraic reconstruction technique (MART) [3][4]. The most important of these, MART, is given by the following iterative update,

$$x_j^{k+1} = x_j^k \left(\frac{y_i}{\langle h_i, x^k \rangle} \right)^{\lambda h_{ij}}, \tag{3}$$

where x_j^k is the jth element of the kth iteration on x, h_i is a vector formed from the ith row of H, $\langle \cdot, \cdot \rangle$ indicates vector inner product, h_{ij} is the i, j^{th} element of H, and λ is a relaxation parameter which controls step size. The general purpose constrained optimization approach is impractical with image sized data sets, and the MART can be very slow. Both are iterative methods, and no closed-form solution has been available.

2. A CLOSED-FORM M.E. APPROXIMATION

In this section we derive a closed-form M.E. approximation by projecting the solution onto the null space of H to formulate an unconstrained optimization problem of lower dimension. A finite series expansion of the entropy expression leads to an algebraic solution to equation (2). This approach will be useful in theoretical analysis of the M.E. problem, and is dramatically faster than MART.

All inverse problems with linear equality constraints, $y = Hx$, have admissible solutions of the form $x = x_o + x_e$, where x_o belongs to the row space of H, (i.e. $x_o \in \Re\{H^T\}$) and x_e is drawn from the right nullspace of H, (i.e. $x_e \in \mathcal{N}\{H^T\}$). x_o is the unique minimum norm solution (known as the least squares solution in the presence of measurement error) given by

$$x_o = H^\dagger y \tag{4}$$

where $\dagger$ indicates pseudo inverse. On the other hand, x_e is different for each optimization criterion that may be chosen, e.g. maximum entropy, least squares, minimum l_p norm, etc. x_e is typically a small perturbation from x_o, but as long as $x_e \in \mathcal{N}\{H^T\}$, x will satisfy the constraint equation. Selecting an optimization criterion in effect determines x_e, and thus determines a particular unique solution for x.

The approach of the proposed algorithm is to first find x_o and then to perturb this solution in the direction of the x_e given by the entropy criterion. x_o may be computed using one of several standard algorithms.

We can decompose H into its range and null spaces using the SVD,

$$H = [U_R | U_N] \Sigma [V_R | V_N]^H \tag{5}$$

where superscript H denotes conjugate transpose and Σ is the diagonal matrix of singular values. U_R and V_R are partitions of left and right singular matrices U and V respectively which correspond to the non-zero singular values of H. Likewise, U_N and V_N contain the singular vectors corresponding to the zero singular values. U_N and V_N span the left and right null spaces of H respectively.

By construction, $\mathcal{N}\{H^T\} = V_N$. Therefore $x_e = V_N z$ for some z. The vector z that leads to a maximum entropy solution of equation (2) is

$$z = \arg\max_z -\sum_{i=1}^{M}(x_{oi} + [V_N z]_i)\ln(x_{oi} + [V_N z]_i), \quad (6)$$

and the M.E. solution for x is simply

$$x_{ME} = x_o + V_N z. \quad (7)$$

Note that this optimization is unconstrained in z. Also, z is length $p = N - rank\{H\} < N$, so expressing x_{ME} in terms of z dramatically reduces both the complexity of the minimization and the number of parameters to be estimated.

Consider the entropy expression from the right hand side of equation (6). Using a finite series expansion approximation yields

$$
\begin{aligned}
\mathcal{E}(x) &= -\sum_i (x_{oi} + x_{ei})\ln(x_{oi} + x_{ei}) \\
&\approx -\sum_i \left(x_{oi}\ln x_{oi} + (1 + \ln x_{oi})x_{ei} + \frac{x_{ei}^2}{x_{oi}}\right) \\
&\approx -(c + r^H x_e + x_e^H B x_e) \\
&\approx -(c + r^H V_N z + z^H V_N^H B V_N z). \quad (8)
\end{aligned}
$$

where the approximation comes from taking the first term of the Taylor series expansion of $\ln(1 + \frac{x_{ei}}{x_{oi}})$. $r = [1 + \ln x_{o1}, \cdots, 1 + \ln x_{on}]^T$, $B = Diag\{[x_{o1}^{-1}, \cdots, x_{on}^{-1}]\}$, and scalar $c = \sum_i x_{oi}\ln x_{oi}$. To maximize entropy we take the derivative with respect to z and set it to zero, which yields

$$V_N^H r + 2V_N^H B V_N z = 0. \quad (9)$$

Therefore, noting that $V_N^H B V_N$ is a full rank square matrix, equations (9) may be solved algebraically for z, which is substituted into equation (7) to yield our final closed form approximation

$$x_{ME} \approx H^\dagger y - \frac{1}{2}V_N(V_N^H B V_N)^{-1}V_N^H r. \quad (10)$$

We note that B and r are direct functions of y through x_o, and V_N is a direct function of H. The required matrix inverse is on a relatively small $p \times p$ matrix. The most significant computation is the singular value decomposition used to form V_N.

2.1. A Fast Closed Form Implementation

If the blur represented by H is due to circular convolution, the singular value decomposition of H becomes trivial using the 2-D FFT, and an extremely fast implementation of equation (10) is possible. For non-circular FIR convolutional blur, the image frame can be extended using a zero fill border with a width equal to the psf region of support. The conventional convolution can then be exactly embedded in a slightly larger circular convolution representation. In this way, most practical restoration problems can utilize the following fast implementation.

For 2-D circular convolution, H is $N \times N$ doubly block circulant and is therefore diagonalized by the 2-D unitary DFT. Thus the SVD of equation (5) is given by $H = \mathcal{F}^H \Sigma \mathcal{F}$, where $\mathcal{F} = \frac{1}{\sqrt[4]{N}}F \otimes F$, F is the 1-D DFT unitary transform matrix, $\otimes$ indicates Kroneker matrix product, and $\Sigma = $

$Diag\{vec\{\text{FFT}_{2D}\{h[m,n]\}\}\}$ [5]. Here the elements of Σ are not ordered by magnitude. Subspace partitioning of H as in (5) yields

$$
V_N = \mathcal{F}P^T, \quad V_R = \mathcal{F}Q^T,
$$

$$
P = \begin{bmatrix}
1 & 0 & & & & & \cdots & & 0 \\
0 & \ddots & & & & & & & \vdots \\
& & 1 & 0 & 0 & & & & \\
& \cdots & 0 & 0 & 1 & 0 & \cdots & & \\
\vdots & & & & & \ddots & & & 0 \\
0 & & & & & \cdots & & 0 & 1
\end{bmatrix}. \quad (11)
$$

P is a $p \times N$ selection matrix formed by deleting all rows from I which correspond to non-zero singular values in Σ. Q is $(N - p) \times N$ and contains the rows of I not in P, such that $P^T P + Q^T Q = I$. Equation (11) alone yields tremendous efficiencies in computing V_N and x_o. $\mathcal{F}$ is known, P and Q are constructed easily by thresholding frequency bin magnitudes in $\text{FFT}_{2D}\{h[m,n]\}$ to identify zero singular values, and $x_o = \mathcal{F}^H Q^T (Q\Sigma Q^T)^{-1}Q\mathcal{F}y$.

Additional computational savings come from recognizing that all products involving V_N^H in equation (10) are simply 2-D FFT's followed by masking out frequency bins corresponding to non-zero singular values. Thus for an arbitrary vector $g = vec\{g[m,n]\}$, $P\mathcal{F}g$ can be interpreted as operator notation for $MASK_P\{\text{FFT}_{2D}\{g[m,n]\}\}$. Likewise, $\mathcal{F}^H P^H g$ corresponds to $\text{IFFT}_{2D}\{MASK_P\{g[m,n]\}\}$. Using FFT's also eliminates the need to store huge $N \times N$ matrices by operating directly on the original $\sqrt{N} \times \sqrt{N}$ images. Without this reduction, the problem is completely intractable for even modest sized images of 256×256 pixels.

The only remaining computational difficulty in equation (10) is the matrix inverse, $(V_N^H B V_N)^{-1}$, which cannot be computed directly in the frequency domain. Instead, we find z with a steepest descent algorithm to find the minimum norm solution to equation (9). The following algorithm operates in the frequency domain to efficiently approximate z in about five iterations using FFTs. No matrices larger than the original images are involved in the computation.

1. $x_o = \mathcal{F}^H Q^T (Q\Sigma Q^T)^{-1}Q\mathcal{F}y$.

2. $z_{k=0} = -\frac{1}{2}P\mathcal{F}[Diag\{x_o\}]\mathcal{F}^H P^T P\mathcal{F}r$.

3. $z_{k+1} = z_k - \mu P\mathcal{F}B\mathcal{F}^H P^T P\mathcal{F}(2B\mathcal{F}^H P^T z_k + r)$.

4. $k = k + 1$. If $k \leq K$ go to 3.

5. $x_{ME} = x_o + \mathcal{F}^H P^T z_k$.

where μ is the iteration step size, and $K \approx 5$ is the desired number of iterations.

2.2. Experimental Results

Figure 1 illustrates an example using the closed-from M.E. approximation. The original Hubble Space Telescope image of planetary nebula NGC 6543 was circularly convolved (i.e. periodic convolution in implemented in the frequency domain) with a low pass filter to produce the output image. No noise was added. Comparing the true maximum entropy solution as produced by MART with the closed form approximation shows striking similarity. Both results recover some of the filament structure in the nebula arms and reduce the blur extent on the caption text. The only apparent difference is an increased low level ringing in Figure 1d, which is

difficult to see in this reproduction. This suggests the new algorithm will be promising for a variety of M.E. image restoration applications. It is noteworthy that the closed form result was computed on a 400 MHz PC in approximately 40 seconds, while the MART computation required more than 24 hours.

3. M.E. WITH NON-UNIFORM SAMPLING

A number of important image restoration and reconstruction applications involve non-uniform or decimated sampling of the observation. An example of this kind of problem is image reconstruction in microwave earth remote sensing, which is of particular interest to the authors. The experiments presented in this section are designed to evaluate M.E. algorithm suitability for this application.

An instrument often used in satellite remote sensing is the scatterometer, which is an orbital radar that transmits microwave energy and measures the return microwave "backscatter." Backscatter has been shown to be useful for ice and vegetation classification. With appropriate reconstruction algorithms, it is possible to form images with much higher resolution than the underlying sample density of the radar system [6]. The area seen by the radar is defined by the radar "footprint," or projection of the antenna aperture function on the ground. By using multiple irregularly spaced overlapping measurements during several orbital passes is possible to form reconstructed images of the ground on a higher resolution regular sample grid. This remote sensing problem is structurally similar to the standard reconstruction from projections problem.

Figure 2 shows the results of some decimated M.E. reconstruction trials using the same planetary nebula image seen in Figure 1(a). Both regular and random decimation of the observed image were studied, with the random decimation by a factor of 36 being representative of the arbitrary sample spacing for a scatterometer antenna ground footprint. The restored image is computed on the original higher resolution undecimated pixel grid.

As seen in Figure 2(a), the M.E. solution does a fair job of reconstructing with regular decimation (though rectangular grid artifacts are obvious). M.E. does not perform as well in the presence of sparse random sampling. Figure 2(b) shows that the M.E. result is extremely corrupted by sampling artifacts, and would be unacceptable as a restoration result. Because spaceborne scatterometer sample spacing is dictated by orbital dynamics and a relatively large antenna footprint, such randomly spaced sampling is unavoidable. We therefore conclude that maximum entropy techniques are not appropriate for these remote sensing applications.

To illustrate that the sparse random decimation example is not an intractable restoration problem, Figure 2(c) presents the result obtained using the same data set as in 2(b), but processed with the Simultaneous Algebraic Reconstruction Technique (SART) [7]. The SART iteration is given by

$$x_j^{k+1} = x_j^k + \frac{\sum_{i=1}^m h_{ij}(y_i - \langle h_i, x^k \rangle)}{\sum_{i=1}^m h_{ij}}. \tag{12}$$

Note the almost total lack of sampling artifacts. SART belongs to a class of related iterative algorithms we have studied where the updates are normalized by the columns of the transfer matrix $\mathcal{H}$. The key feature of this algorithm which reduces sampling artifacts is the normalizing denominator term, $\sum_{i=1}^m h_{ij}$. We have shown that a variety of algorithms with this term are artifact suppressing, and thus well suited for microwave earth remote sensing imaging [8].

4. CONCLUSIONS

Maximum entropy image restoration is an effective approach which yields surprisingly good results given the simplicity of the optimization criterion. The M.E. criterion includes no spatial correlation metric (i.e. there are no cross pixel terms in the criterion), yet M.E. restorations are known to be edge preserving and to perform well when flat intensity regions are involved. Unfortunately the best existing algorithm, MART, is notorious for slow iterative convergence. The new fast closed form M.E. approximation solves this problem, and makes quick restorations possible even for very large images. This may make M.E. a more attractive option for future use in a variety of applications

We have also shown that M.E. reconstruction is not appropriate for sparsely sampled images. This suggests that a solution which combines the advantages of M.E. and artifact suppressing algorithms like SART may be very useful. We are currently studying a new column normalized algorithm with multiplicative updates like MART and plan to report on our findings in the future.

5. REFERENCES

[1] S. F. Gull and J. Skilling, "Maximum entropy method in image processing," *Proceedings of the IEEE*, vol. 131, pt. F, no. 6, pp. 646–659, 1984.

[2] E. Jaynes, "On the rational of maximum-entropy methods," *Proceedings of the IEEE*, pp. 936–952, 1982.

[3] Y. Censor, "Finite series-expansion reconstruction methods," *Proceedings of the IEEE*, vol. 71, pp. 409–418, 1983.

[4] T. Elfving, "On some methods for entropy maximization and matrix scaling," *Linear Algebra and its Applications*, vol. 34, pp. 321–329, 1980.

[5] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, 1989.

[6] D. Long and P. Harding, "Vegetation studies of the amazon basin using enhanced resolution seasat scatterometer data," *IEEE Trans. Geosc. Remote Sens.*, vol. 32, pp. 449–460, Mar. 1994.

[7] A.H. Andersen and A.C. Kak, "Simultaneous algebraic reconstruction technique (sart): A superior implementation of the art algorithm," *Ultrasonic Imaging*, vol. 6, pp. 81–94, 1984.

[8] Matthew Willis, "Algebraic reconstruction algorithms for remote sensing image enhancement," M.S. thesis, Brigham Young University, 2000.

Fig. 2. Examples of reconstructed images. a) Regularly decimated by a factor of 6 in both the x and y directions, and reconstructed using MART. b) Randomly decimated by a factor of 36 (i.e. one pixel in 36 retained) and reconstructed with MART. c) Randomly decimated by 36 and reconstructed with SART.

Fig. 1. Closed form maximum entropy results. a) Original 256 × 256 pixel image. b) Observed image, blurred with a low pass filter with cutoff frequency of $\frac{\pi}{2}$. c) M.E. reconstruction using MART. d) Closed form approximation of M.E.

92