# A New Look at Maximum Entropy Image Reconstruction

Matthew Willis, Brian D. Jeffs and David G. Long*
Department of Electrical and Computer Engineering
Brigham Young University, Provo, UT 84602
willism@et.byu.edu bjeffs@ee.byu.edu long@ee.byu.edu

## Abstract

*This paper presents new insights into the maximum entropy (ME) method of image restoration. It is shown that when a specific image prior probability pdf model is chosen for Bayesian MAP restoration, the resulting solution is identical to the maximum entropy result. This relationship provides a new means of evaluating the theoretical foundations of maximum entropy, and may assist in determining what class of images are best suited for ME processing. Also, a new non-iterative, closed-form approximation to the ME solution is developed. This result can reduce computational demands compared to conventional iterative algorithms. An example of the closed form restoration is presented.*

## 1. Introduction

Maximum entropy (ME) image restoration has been widely used with significant success for many years. It remains the method of choice for radio astronomy image restoration applications. Though usually expressed as a deterministic constrained convex optimization problem, proponents of ME restoration have used information theoretic and quantum physics arguments to justify its use in a wide variety of applications [5][7].

A frequently used observation model for maximum entropy restoration is

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

where $y : M \times 1$ and $x : N \times 1$, are vectors formed by column scanning the corresponding 2-D images, and $H$ is the doubly block Toeplitz convolution matrix corresponding to the 2-D point spread function. $\eta$ is a random vector, with joint probability density function (pdf), $p_\eta(\eta)$, which represents measurement uncertainty or additive observation noise. For simplicity we will assume each noise pixel has identical variance, $\sigma_\eta^2$.

Many ME restoration methods employ an equality constraint by assuming $\sigma_\eta^2 = 0$ [2] [4]. In this case it must also be assumed that $H$ does not have full column rank, so that given $y$, the solution for $x$ is not unique unless some optimization criterion (such as maximum entropy) is employed. The corresponding equality constrained ME optimization problem for image restoration is given by

$$\begin{aligned} \hat{x}_{ME1} &= \arg\max_{x \geq 0} -\sum_{i=1}^{M} x_i \ln x_i, \\ &\quad \text{such that } y = Hx. \end{aligned} \tag{2}$$

We also consider the case of inconsistent measurements, i.e. when $\sigma_\eta^2 > 0$, which can be addressed with an $l_2$ vector norm constraint on the entropy expression [6]:

$$\begin{aligned} \hat{x}_{ME2} &= \arg\max_{x \geq 0} -\sum_{i=1}^{M} x_i \ln x_i, \\ &\quad \text{such that } \| y - Hx \|^2 = 2\sigma_\eta^2. \end{aligned} \tag{3}$$

A variety of iterative algorithms are available to solve ME expressions (2) and (3) [2][6], but no practical direct form solutions are known.

## 2. Relationship Between Maximum Entropy and Bayesian Image Restoration

In this section we will show a relationship between the maximum entropy solutions of equations (3) and (2) and the well known and widely used maximum *a posteriori* (MAP) Bayesian image restoration technique. It will be shown that for a particular choice of the image prior probability density model, the MAP solution is equivalent to maximum entropy. Since MAP restoration is a well known method founded in classical estimation theory, knowing the limiting conditions under which the two solutions match will give us insight into properties of maximum entropy methods. In

[0]The authors acknowledge the contribution of Cornell Bean to this work.



0-7803-5700-0/99/$10.00©1999 IEEE

particular, MAP estimation incorporates prior information about the solution, $x$, in the form of a prior probability density $p_x(x)$. Knowing which $p_x(x)$ yields the ME solution will suggest what kinds of implicit modeling assumptions are being made with regard to $x$.

An simple example of the utility of establishing such relationships is found in least squares optimization. Least squares is often used as a deterministic approach, without regard to signal statistics, because of its ease of application and the quality of results obtained. However, the MAP estimator for images with uniformly distributed intensities, and which are corrupted by white Gaussian noise, has exactly the same form as least squares with bounded intensities. This knowledge provides insight into the power of the method, and guides practitioners in deciding when least squares may be appropriate based on signal statistics (e.g. when the truth image is actually uniformly distributed).

It must be pointed out that a complete image model paradigm shift is needed in order to use estimation theoretic methods for restoration. For maximum entropy, the image pixel intensity values themselves are modeled as probabilities (e.g. $x_i$ is the probability that a single detected photon strikes pixel $i$.) On the other hand, in the conventional signals and systems view, all images (including the true image, the observed image, and noise) are modeled as *realizations* drawn from random fields. The statistical properties of an image are completely described by its probability density function, which is distinct from the image realization itself. Restoration is thus a statistical estimation problem, i.e. finding the $x$ which best optimizes some statistical measure. In this context the maximum entropy optimization of equation (3) has no obvious relationship to any well known statistical estimator.

The MAP estimator solution can be expressed as

$$\begin{aligned} \hat{x}_{MAP} &= \arg\max_x \; p_{x|y}(x|y), \\ &= \arg\max_x \; \ln\{p_{y|x}(y|x)p_x(x)\}, \end{aligned} \tag{4}$$

where $p_{y|x}(y|x)$ is the conditional probability density for observed image $y$ and $p_x(x)$ contains all of our knowledge about the expected form of the true image. It can be shown that $p_{y|x}(y|x)$ is simply a mean-shifted version of the noise distribution, $p_\eta(\eta)$. The current state of the art is to use Markov random field models based on a variety of Gibbs distributions for $p_x(x)$ in order to represent ones prior notions of the correlation, texture, and edge activity structure expected in $x$ [1, 3].

We will assume that the noise is distributed i.i.d. zero mean Gaussian with variance $\sigma_\eta^2$. Thus $p_{y|x}(y|x) = p_\eta(y - Hx)$, as given by

$$p_{y|x}(y|x) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}^{\,M}} \exp\left\{-\frac{1}{2\sigma_\eta^2} \parallel y - Hx \parallel^2\right\}, \tag{5}$$

where $\| \cdot \|^2$ denotes $l_2$ vector norm squared.

We now make a less obvious assumption: that pixels $x_i$ are distributed i.i.d. with marginal and joint image prior pdf's given respectively by

$$p_x(x_i) = \frac{1}{Z} x_i^{-x_i} u(x_i), \text{ and} \tag{6}$$

$$p_x(x) = \frac{1}{Z^M} \prod_{i=1}^{M} x_i^{-x_i} u(x_i), \tag{7}$$

where $Z = \int_0^\infty x_i^{-x_i} dx \approx 2.0$ is the constant valued partition function which insures $p_x(x_i)$ integrates to 1, and $u(\cdot)$ is the unit step function. Figure 1 illustrates $p_x(x_i)$. Note that this is a one sided, highly skewed distribution with a modal peak near zero.

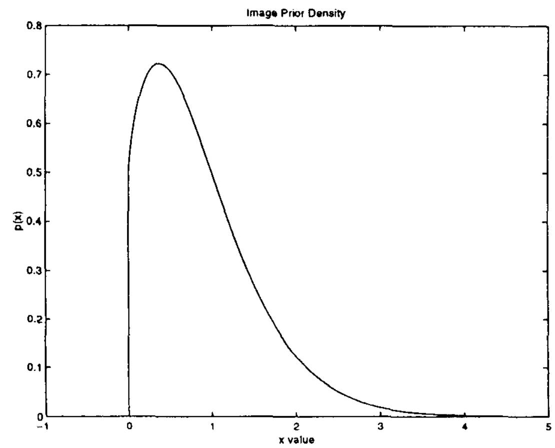



Figure 1. Marginal image prior density, $p_x(x_i) = \frac{1}{Z} x_i^{-x_i} u(x_i)$. When $x$, is distributed i.i.d. with this density, and the noise is Gaussian, the MAP solution is identical to the $l_2$ norm constrained maximum entropy solution of equation (3).

Substituting equations (5) and (7) into (4) leads to

$$\hat{x}_{MAP} = \arg\max_{x \geq 0} \left\{ -\frac{1}{2\sigma_\eta^2} \parallel y - Hx \parallel^2 + \ln \prod_{i=1}^{M} x_i^{-x_i} \right\}, \tag{8}$$

$$\hat{x}_{MAP} = \arg\max_{x \geq 0} \left\{ -\sum_{i=1}^{M} x_i \ln x_i + \lambda \parallel y - Hx \parallel^2 \right\}, \tag{9}$$

where $\lambda = -\frac{1}{2\sigma_\eta^2}$ and additive constants have been dropped due to the maximum operator.

Equation (9) is readily seen to be the Lagrangian expression corresponding to the constrained optimization of equation (3), with $\lambda$ the Lagrange multiplier. The important consequence of this observation is that the solutions are equal, i.e. $\hat{x}_{MAP} = \hat{x}_{ME2}$. This is noteworthy because the solutions were derived using entirely different image models, from information theoretic vs. estimation theoretic perspectives, and with different assumptions about the image signals. The equality constraint ME solution is also equivalent to a MAP solution. It can be shown that in the limit as $\sigma^2 \rightarrow 0$ equation (9) corresponds to an equality constraint identical to the $\hat{X}_{ME1}$ solution of equation (2).

A key conclusion to be drawn from the equivalence of the ME and MAP solutions relates to the form of $p_x(x_i)$ shown in Figure 1. We know of no image class where this specific model can be justified by physical processes or by empirical estimation of the pdf using histogram methods. One may observe however that $p_x(x_i)$ is one sided, heavy tailed, and has a modal peak in the low intensity values. It may be argued that these characteristics are consistent with some photographic images which have large darker background regions surrounding a brighter high contrast object of interest. The overall shape of $p_x(x_i)$ is grossly similar to a non-central Chi-squared distribution, which could arise from a square law detector (intensity proportional to power) in a linear observation environment that produces Gaussian random variates due to central limit theorem effects.

These observations suggest one should be cautious about using maximum entropy, at least from the estimation theoretic, signals and systems point of view. Typical images of interest may not have the distributions we have implicitly assumed. Further, most images have significant interpixel correlation and are thus not i.i.d. as with the $p_x(x)$ of equation 7. In these cases the ME solution will not be optimal in the Bayesian sense. Using the ME approach is equivalent to assuming the "unlikely" image prior density, $p_x(x_i) = \frac{1}{Z} x_i^{-x_i} u(x_i)$, which is difficult to justify for many problems. This reinforces our concern that the ME approach of interpreting the image itself as a probability *distribution* in order to force-fit the information theoretic entropy measure onto the image restoration problem is not well supported, unless the true image pdf is similar to equation (7).

On the other hand, the ME criterion has enjoyed much success due to a number of advantages. ME restorations have been found to preserve edge and point-like image detail [5]. Equations (2) and (3) define convex deterministic optimizations over convex constraints, a formulation which has significant computational advantages as compared with many MAP algorithms. Efficient iterative algorithms exist to solve these ME problems (e.g. multiplicative algebraic reconstruction technique, MART [2]). We find ME restoration to be a powerful image processing tool, but prefer to view it as a deterministic constrained optimization method with an arbitrary, though effective optimization metric. This metric was developed by *analogy* with information theory, and has been shown to perform well for a large class of images of interest. Despite a controversial theoretical foundation, we believe that ME methods stand on their own merits based on performance.

## 3. A Closed-Form Maximum Entropy Approximation

The ME 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) [2]. The former approach is impractical with image sized data sets, and the later can be very slow. Both are iterative methods, and no closed-form solution is available. In this section we derive an approximate closed form result for the equality constraint maximum entropy problem of equation (2). This approach will be useful in theoretical analysis of the ME problem, and in some cases will provide a result with lower computational demands 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

$$x_o = H^\dagger y \tag{10}$$

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 $x_e$ to yield 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, e.g. singular value decomposition based pseudoinverse, iterative least squares, or additive algebraic reconstruction technique (ART).

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

$$H = \left[ [U_R|U_N] \sum^{\frac{1}{2}} [V_R|V_N]^H \right] \tag{11}$$

where superscript $^H$ denotes conjugate transpose. $U_R$ 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 will leads to a maximum entropy solution of equation (2) is

$$z_{ME1} = \arg\max_z - \sum_{i=1}^{M} (x_{oi} + [V_N\, z]_i) \ln(x_{oi} + [V_N\, z]_i), \tag{12}$$

and the ME solution for $x$ is simply

$$x_{ME1} = x_o + V_N\, z_{ME1}. \tag{13}$$

Note that this optimization is unconstrained in $z$. Also, $z$ is length $P = N - rank\{H\} \ll N$, so expressing $x_{ME1}$ 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 (12), which for $x_e = V_N\, z$ is $\mathcal{E}(x) = -\sum_{i=1}^{M} x_i \ln x_i$. Using a finite series expansion approximation yields

$$\begin{aligned} \mathcal{E}(x) &= -\sum_i (x_{oi} + x_{ei}) \ln(x_{oi} + x_{ei}) \\ &= -\sum_i (x_{oi} + x_{ei}) \ln x_{oi}(1 + \frac{x_{ei}}{x_{oi}}) \\ &= -\sum_i (x_{oi} + x_{ei}) \left[\ln x_{oi} + \ln(1 + \frac{x_{ei}}{x_{oi}})\right] \\ &\approx -\sum_i (x_{oi} + x_{ei}) \left(\ln x_{oi} + \frac{x_{ei}}{x_{oi}}\right) \end{aligned} \tag{14}$$

where the approximation comes from taking the first term of the Taylor series expansion of $\ln(1 + \frac{x_{ei}}{x_{oi}})$. Expanding the terms in the summation gives

$$\begin{aligned} \mathcal{E}(x) &\approx -\sum_i \left(x_{oi} \ln x_{oi} + (1 + \ln x_{oi}) x_{ei} + \frac{x_{ei}^2}{x_{oi}}\right) \\ &\approx -(c + r^H x_e + x_e^H B x_e) \end{aligned} \tag{15}$$

where $r = [1 + \ln x_{o1}, \cdots, 1 + \ln x_{on}]^H$, $B = diag(x_{o1}^{-1}, \cdots, x_{on}^{-1})$, and scalar $c = \sum_i x_{oi} \ln x_{oi}$ (which, incidentally, is the negative entropy of $x_{oi}$). Substituting $V_N\, z$ for $x_e$, we obtain

$$\begin{aligned} \mathcal{E}(x) &\approx -(c + r^H V_N z + z^H V_N^H B V_N z) \\ &\approx -(c + S z + z^H D z) \end{aligned} \tag{16}$$

where $D = V_N^H B V_N$ and $S = r^H V_N$. To maximize entropy we take the derivative with respect to $z$ and set it to zero,

$$S^H + 2Dz = 0. \tag{17}$$

Therefore, noting that $D$ is a full rank square matrix, the solution to equation (12) is

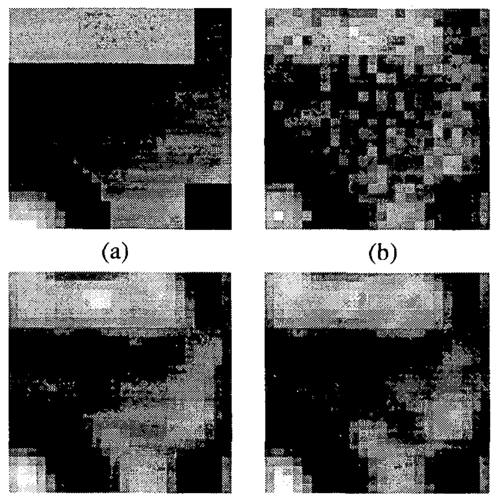
$$\begin{aligned} z_{ME1} &= -\frac{1}{2} D^{-1} S^H \\ &= -\frac{1}{2} (V_N^H B V_N)^{-1} (r^H V_N)^H. \end{aligned} \tag{18}$$

Substituting into equation (13) yields our final closed form approximation

$$x_{ME1} \approx H^\dagger y - \frac{1}{2} V_N (V_N^H B V_N)^{-1} (r^H V_N)^H. \tag{19}$$

We note that $B$ and $r$ are direct closed form 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 $V_N$. However if the point-spread-function is known *a priori*, or is used for a series of images, $V_N$ can be computed in advance so the actual image restoration computation dependent on data $y$ is small.

Figure 2 illustrates an example using the closed-from ME approximation. The original image was circularly convolved with a low pass filter, and then noise was added to produce the output image. Comparing the true maximum entropy solution as produced by MART with the approximation shows how similar the two are. This suggests the algorithm will be promising for larger scale ME image restoration applications.



(a) (b) (c) (d)

Figure 2. Example results from the closed form maximum entropy algorithm of equation (19). a) Original $27 \times 27$ pixel true image. b) Observed image, $13 \times 13$ pixels. Blurred with 5 by 5 pixel raised cosine psf, and decimated by 2 in each axis. c) True maximum entropy reconstruction. $27 \times 27$ pixels. d) Closed form approximation of maximum entropy. Note similarity of solutions in (d) and (c) confirms accuracy of the closed form approximation.

## 3.1 Conclusions and Future Work

In this paper we have shown a relationship between maximum entropy and Bayesian MAP image restorations under the assumption of a particular image prior pdf. This provides insight into what class of images are best suited to ME restoration by suggesting a pdf model for the random field they should be drawn from. An obvious next step is to evaluate an ensemble of images that are typically processed using ME (e.g. radio astronomical images) and compare their histograms with $p_x(x_i)$ of Figure 1. A close match would indicate that ME restoration has remained popular in these application areas, despite the availability of many other methods, because unknown to the practitioners the underlying implicit image prior pdf model was a good fit.

A closed form approximation to the ME solution was developed 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 led to an algebraic solution to this optimization. The new approximation can be much faster (particularly for a set of observations with common $H$) and will provide insight into the structure of ME results.

Equation (19) was derived using a first order approximation for the logarithm. Using a second order term for the log expansion would give a more accurate approximation. The approximate entropy expression would then be

$$\mathcal{E}(x) \approx -\sum_i (x_{oi} + x_{ei}) \left( \ln x_{oi} + \frac{x_{ei}}{x_{oi}} - \frac{1}{2} \left( \frac{x_{ei}}{x_{oi}} \right)^2 \right) \tag{20}$$

where now the second order Taylor series for the logarithm has been used. We are currently developing a closed form optimization for equation (20).

## References


[1] C. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. on Image Processing*, 2(3):296–310, July 1993.

[2] Y. Censor. Finite series-expansion reconstruction methods. *Proceedings of the IEEE*, 71:409–418, 1983.

[3] R. Chellappa and A. Jains, editors. *Markov random fields: theory and applications*. Academic Press, San Diego, 1991.

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

[5] S. F. Gull and J. Skilling. Maximum entropy method in image processing. *Proceedings of the IEEE*, 131, pt. F(6):646–659, 1984.

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

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