Restoration of Blurred Star Field Images by Maximally Sparse Optimization

Brian D. Jeffs, *Member, IEEE,* and Metin Gunsay

Abstract—**In this paper we address the problem of removing blur from, or sharpening, astronomical star field intensity images. A new approach to image restoration is introduced which recovers image detail using a constrained optimization theoretic approach. Ideal star images may be modeled as a few point sources in a uniform background. It is therefore argued that a direct measure of image sparseness is the appropriate optimization criterion for deconvolving the image blurring function. A sparseness criterion based on the l_p is presented and candidate algorithms for solving the ensuing nonlinear constrained optimization problem are presented and reviewed. Synthetic and actual star image reconstruction examples are presented which demonstrate the method's superior performance as compared with several standard image deconvolution methods.**

I. Introduction

BLUR in long exposure astronomical star images may be due to atmospheric turbulence, misfocus, poor telescope tracking, finite aperture size, or other optical distortion effects. This blurring operation in incoherent intensity images may be modeled as the two dimensional convolution of a point spread function (usually low pass in spatial frequency) with the uncorrupted true image. If the point spread function is known, or may be estimated, then image resolution may be improved by any of many published deconvolution techniques [1]–[8]. However, due to the typically low pass nature of the blurring function, observation noise, and the ill posed nature of the problem, a unique "best" solution is not possible in general, and the type of solution obtained from any given algorithm is highly dependent on its underlying (explicit or implicit) objective function.

We postulate that ideal star images are inherently sparse in nature, that is they are dominated by a constant flat field background intensity level with a small percentage of image pixels containing star intensity information. This image model justifies using a maximally sparse optimization criterion in the reconstruction algorithm, which in turn enables dramatic improvement in resolution. We adopt the following linearized image degradation model

$$\underline{b} = \boldsymbol{H}\underline{x} = \underline{\eta} \tag{1.1}$$

where $\underline{b}$ is the observation vector obtained by row scanning the sampled 2-D degraded image, $\underline{x}$ is the uncorrupted image vector, $\boldsymbol{H}$ is the block Toeplitz matrix representing the row scanned convolutional blur point spread function, and $\underline{\eta}$ is the additive observation noise vector. $\boldsymbol{H}$ is assumed known and can be modeled by sampling a standard atmospheric blurring function [9], or estimated from nearby isolated stars.

The restoration problem is cast as one of linear inequality constrained nonlinear minimization

$$\min_{\underline{x}} g(\underline{x}) \qquad \text{such that} \qquad |\boldsymbol{H}\underline{x} - \underline{b}| \leq \underline{\epsilon}, \underline{x} \geq \underline{0} \tag{1.2}$$

where the constraint vector $\underline{\epsilon}$ represents our observation uncertainty due to additive noise $\underline{\eta}$ and possible error in our knowledge of the point spread function. The inequality ($\leq$) indicates the element by element relationship between vectors $|\boldsymbol{H}\underline{x} - \underline{b}|$ and $\underline{\epsilon}$. The solution, $\underline{x}$, is constrained to be nonnegative since we are dealing with intensity images from an incoherent imaging system.

There are in general an infinite number of admissible solutions which are consistent with the observed degraded image and satisfy the constraints of (1.2). The choice of objective function, $g(\underline{x})$, is key to controlling the form of the solution image, $\underline{x}$. The more common objectives, l_2 vector norm, entropy, etc., yield unacceptably "smooth" results which often do not achieve maximum improvement in image resolution [10]. l_2 minimization, as implemented in algorithms such as the Algebraic Reconstruction Technique (ART) [3], finds the minimum length vector $\underline{x}$ that satisfies the inequality constraints, and in doing so typically distributes energy throughout most of the image pixels. Though maximum entropy reconstruction has been proposed for astronomical image restoration [4], [5], the underlying assumption in entropy maximization is that the solution image should be as uniformly distributed as possible, consistent with the observed data. This too can have a low-pass filter effect, distributing energy among many pixels. Representing isolated stars requires significant high spatial frequency content and implies that the restoration process should localize image energy into sharp high resolution peaks, rather than distributing energy. Some authors have presented techniques which do emphasize sharpness, or high frequencies, but these methods often do not take full advantage of the truly sparse nature of star fields [2], [6], [7], [8]. Section V contains experimental results which demonstrate the improved resolution of the proposed algorithm as compared to minimum norm, maximum entropy, and CLEAN [6] restoration for star field images.

Our prior knowledge that the desired image is sparse suggests that the appropriate objective function is a direct measure of solution sparseness. The restoration problem may then be

Manuscript received March 5, 1992; revised August 17, 1992. This work was supported in part by the National Science Foundation under Grant MIP-9110187. The associate editor coordinating the review of this paper and approving it for publication was Dr. M. Ibrahim Sezan.

The authors are with the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602.

IEEE Log Number 9206906.

restated as, "find the solution $\underline{x}$ which has the fewest possible nonzero elements and satisfies $|\boldsymbol{H}\underline{x} - \underline{b}| \leq \underline{\epsilon}$." To this end, we propose an objective, $g(\underline{x}) = \Sigma_{i=1}^{N} |x_i|^p$, which is related to the l_p , and which will be shown to be an excellent sparseness metric when $0 < p < 1$. Equation (1.2) may then be expressed as

$$\min_{\underline{x}} g(\underline{x}) = \sum_{i=1}^{N} (x_i)^p$$

such that

$$|\boldsymbol{H}\underline{x} - \underline{b}| \leq \underline{\epsilon},$$

$$0 < p < 1, \qquad \underline{x} \geq 0 \tag{1.3}$$

In the following section we justify this choice of $g(\underline{x})$ as a sparseness measure and develop the theory for l_p quasi-norm sparse optimization. In Section III we discuss algorithms for approximate and global solutions to (1.3). Section IV contains a discussion comparing the underlying objective functions for the proposed approach and the well-known CLEAN algorithm. It is shown that CLEAN is suboptimal for star field images. Section V presents experimental results of deblurring both synthetic and actual star field images.

II. Maximally Sparse Star Reconstruction With l_p Quasinorms

This section develops an argument for using l_p s as the objective function in optimal star deblurring. Two different classes of optimization problems are considered: Deterministic star deblurring, and an estimation theoretic approach. For the deterministic interpretation we begin with the assumption that our underlying true image is maximally sparse as defined above and that the restored image should maximize some direct measure of sparseness while controlling reconstruction error. Applying these assumptions leads to a deterministic nonlinear optimization formulation. In the probabilistic approach we view both the true star image and observation noise as random fields and use Bayesian estimation theory to recover the star image. A class of probability density functions is identified which provide a good model for the prior distribution of the true star image. Maximum *a posteriori* estimators for this image under two different noise distributions are developed. It is notable that both deterministic and estimation theoretic interpretations of star deblurring are shown to imply the suitability of l_p quasi-norm optimization and lead to a problem of the form of (1.3).

As with all image models, we cannot expect that the maximally sparse assumption is accurate in all possible blurred star images. Although we have not encountered the situation in practice, it is possible that a true star could be lost in the reconstruction due to the sparseness criterion. In such a case, however, the solution with fewer stars would be as consistent with the observed data as the true image, and the sparseness criterion resolves the ambiguity. When equally valid solutions exist (i.e., each satisfying the inequality constraint of (1.3) containing different numbers of isolated points, the sparseness criterion favors the solution with fewest stars since there is no evidence to the contrary.

A. Deterministic Optimization Formulation

The most obvious measure of image sparseness is a simple count of the nonzero pixels. Since we have modeled stars as point sources, this corresponds to a count of the number of stars resolved in the restored image. This may be accomplished with an objective, $f(x)$, based on the indicator function:

$$f(\underline{x}) = \sum_{i=1}^{N} 1(x_i),$$

$$1(x_i) = \begin{cases} 1, & x_i \neq 0 \\ 0, & x_i = 0 \end{cases}. \tag{2.1}$$

$f(\underline{x})$ is not well suited as an objective function in an optimization algorithm and does not permit any control over the degree of sparseness desired [11]. It is discontinuous in the regions of interest (where any x_i goes to zero), and has a gradient of zero everywhere else, thus making it useless for gradient descent techniques. $g(\underline{x})$ as defined above achieves an equivalent measure of sparseness without some of these difficulties. To demonstrate this, consider the unit ball surfaces in R^2 space for the $||\underline{x}||_{l_p} = (\Sigma_{i=1}^{N} |x_i|^p)^{1/p}$ as shown in Fig. 1 for different values of p. For $p \geq 1$ we have the conventional l_p vector norm. The linear constraints in (1.3) form a convex set, and it is well known that within convex constraints, a local minimum of $||\underline{x}||_{l_p}, p \geq 1$, is a global optimum. Many efficient gradient descent algorithms exist for solving such problems [12]. Of particular interest are the cases for values of $p = 1, 2$, and ∞, corresponding to linear, least squares, and minimax objective functions, which form the basis of many widely used optimization procedures. However these methods do not achieve sparse results for an underdetermined problem of the form of (1.2).

For $0 < p < 1, ||\underline{x}||_{l_p}$ is only a . Over $R^N, ||\underline{x}||_{l_p}$ is neither convex nor concave, contains many strong local minima and presents a difficult optimization problem. Large values of p result in smooth solutions, however, as $p \to 0$ the solutions tend to become more "spiky," or sparse [13]. The reason for this can be seen in Fig. 1. As $p \to 0$, the unit ball curves approach the x_1, x_2 axes, which is exactly where the unit ball surface lies for the indicator function of (2.1). We may therefore identify minimum order optimization as a special case of generalized l_p optimization. $g(\underline{x}) = (||\underline{x}||_{l_p})^p$, thus

$$\lim_{p \to 0} g(\underline{x}) = \lim_{p \to 0} \sum_{i=1}^{N} |x_i|^p = \sum_{i=1}^{N} 1(x_i) = f(\underline{x}). \tag{2.2}$$

This suggests that we may use (at least in the limiting case) $g(\underline{x})$ from (1.3) for sparse optimization. We have shown elsewhere that if the set of feasible solutions to the constraint equation of (1.3) are bounded in magnitude, then there is a finite $p_0 > 0$ such that for all $0 < p \leq p_0$ any solution to (1.3) is in fact maximally sparse [11]. p need not approach zero to achieve sparse results. The utility of this observation is that for p finite, $g(\underline{x})$ eliminates some of the handicaps of using $f(\underline{x})$ in an optimization algorithm while still yielding

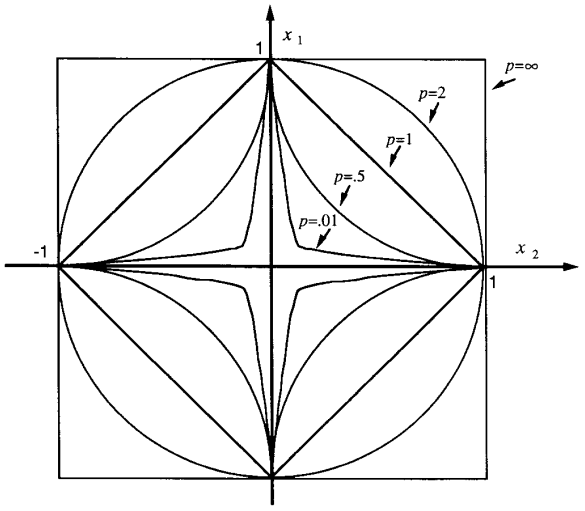


Fig. 1. Unit balls of the l_p norm for various p. Note that as p approaches 0, the unit ball approaches the axes.

a maximally sparse solution. $g(\underline{x})$ is continuous everywhere and differentiable except at the coordinate axes where some $x_i = 0$. We may also adjust the desired sparseness of the solution by varying p in the range $p_0 \leq p \leq 1$.

B. Estimation Theoretic Formulation

In this section a probabilistic model is adopted with an assumed prior distribution on $\underline{x}$ which is well suited to describe a sparse star image as a random field. Bayesian estimation techniques are then applied to develop the maximum *a posteriori* estimate for a given noise distribution. It is then shown that this leads to an estimator which is closely related to the deterministic optimization approach developed above.

The generalized p Gaussian (gpG) probability density function, also known as the Box—Tiao distribution, defines a family of distributions which can be used to characterize non-Gaussian sample data, and in particular sparse or spiky data produced by a heavy tailed distribution. The gpG densities were introduced by Subbotin [14] in 1923, and used by Miller and Thomas [15] in 1972 for modeling non-Gaussian noise in detection theory. McDonald has shown how this parametric family includes many well known symmetric probability density functions [16]. The univariate density is defined for shape parameter p, mean μ, and variance σ^2 as

$$f_z(z) = gpG(\mu, \sigma^2) = \frac{p}{2\Gamma(1/p)\gamma\sigma} e^{-(|z-\mu|/\gamma\sigma)^p}$$
$$\gamma = \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{1/2} \tag{2.3}$$

where $\Gamma()$ is the gamma function. For $p = 2$ and $p = 1$ this yields the familiar Gaussian and double exponential distributions respectively, and as $p \to \infty$ we have a uniform distribution.

By adjusting the shape parameter, p, we may produce a close match to the statistics of a remarkably wide range of sampled data distributions [13], [16]. For the problem at hand, the primary range of interest is $0 < p < 1$. Fig. 2 shows a comparison between synthesized independent Gaussian and gpG data (with $p = 0.3$). The data are presented as 2-D random fields (images) with the pixel intensity corresponding to the absolute value of the random sample. Note that the gpG data is clearly a better model for an unblurred star image and that for this small p, the sequence is much more "spiky," containing primarily small values with a few large outlying spikes. This sparse data is consistent with our view of how noisy samples from a maximally sparse source should appear. Indeed, several authors report excellent data modelling using low order gpG distributions [13], [17]–[19].

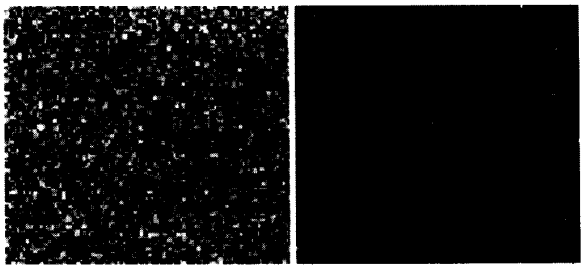


Fig. 2. Comparison of gaussian and generalized p-Gaussian data (a) One-sided gaussian random field, $\sigma = 1$. (b) One sided gpG random field, $p = 0.3$, $\sigma = 1$. Note the similarity of (b) to a star field image.

We shall adopt a stochastic model for unblurred star images in which the pixel intensity prior distribution is independent and identically distributed, one sided gpG with $0 < p < 1$. Outlying samples correspond to single pixel stars. Statistical independence from pixel to pixel is appropriate for this model since stars are randomly placed point sources and spatial correlations in the desired true image are negligible. A minor modification of the gpG function is required to accommodate one sided data needed to model intensity images with only nonnegative values

$$f_x(x) = \begin{cases} \dfrac{p}{\Gamma(1/p)\gamma\sigma} e^{-(x/\gamma\sigma)^p}, & x \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{2.4}$$

We now formulate maximum *a posteriori* (MAP) estimates of the N element image vector $\underline{x}$, given the M element observed image vector $\underline{b}$ for two different noise distributions (see (1.1) for the image observation model).

Case 1: Uniform iid noise

$$f_x(\underline{x}) = gpG(0, \sigma_x^2)$$

i.e., $\underline{x}$ is distributed iid, zero mean, one sided gpG, $0 < p < 1$

$$f_\eta(\eta) = U(-\epsilon, \epsilon) = \begin{cases} \dfrac{1}{2\epsilon}, & -\epsilon \leq \eta \leq \epsilon \\ 0, & \text{otherwise} \end{cases} \tag{2.5}$$

which yields the conditional density

$$f_b(\underline{b}|\underline{x}) = f_\eta(\underline{b} - \boldsymbol{H}\underline{x})$$
$$= \begin{cases} \left(\dfrac{1}{2\epsilon}\right)^M & |\underline{b} - \boldsymbol{H}\underline{x}| \leq \underline{\epsilon}, \\ & \text{(where } \underline{\epsilon} = (\epsilon, \epsilon, \cdots \epsilon)^t\text{).} \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

The MAP estimate of $\underline{x}$ is then

$$\hat{\underline{x}}_{MAP} = \max_{\underline{x}} f_x(\underline{x}|\underline{b}) = \max_{\underline{x}} f_b(\underline{b}|\underline{x}) f_x(\underline{x}) = \max_{\underline{x}} \begin{cases} \left(\frac{1}{2\epsilon}\right)^M \left[\frac{p}{2\Gamma(1/p)\gamma\sigma}\right]^N e^{-\Sigma_i (|x_i|/\gamma\sigma)^p} \\ \quad \text{if } |\underline{b} - \boldsymbol{H}\underline{x}| \le \underline{\epsilon} \text{ and } x_i \ge 0 \\ 0, \text{otherwise.} \end{cases} \tag{2.7}$$

Ignoring constants and taking the logarithm

$$\hat{\underline{x}}_{\text{MAP}} = \min_{\underline{x}} \sum_i |x_i|^p \quad \text{such that} \quad |\underline{b} - \boldsymbol{H}\underline{x}| \le \underline{\epsilon} \text{ and } x_i \ge 0 \tag{2.8}$$

which corresponds exactly with the l_p optimization problem of (1.3).

It may be argued, however, that uniformly distributed noise is atypical. We therefore consider the more realistic case of Gaussian noise.

Case 2: Gaussian noise

$$f_x(\underline{x}) = gpG(0, \sigma_x^2)$$
$$f_\eta(\eta) = N(0, \sigma_\eta^2), \text{iid, white zero mean Gaussian}$$
$$f_b(\underline{b}|\underline{x}) = \frac{1}{(\sqrt{2\pi}\sigma_\eta)^M} \exp - \frac{(\underline{b} - \boldsymbol{H}\underline{x})^T (\underline{b} - \boldsymbol{H}\underline{x})}{2\sigma_\eta^2} \tag{2.9}$$

which leads to the following MAP estimate of $\underline{x}$

$$\hat{\underline{x}}_{\text{MAP}} = \max_{\underline{x}} \left\{ \exp - \frac{||\underline{b} - \boldsymbol{H}\underline{x}||^2}{2\sigma_\eta^2} \exp - \sum_i \left(\frac{|x_i|}{\gamma\sigma_x}\right)^p \right\}, \quad x_i \ge 0$$
$$= \min_{\underline{x}} \left\{ \frac{||\underline{b} - \boldsymbol{H}\underline{x}||^2}{2\sigma_\eta^2} + \left(\frac{1}{\gamma\sigma_x}\right)^p \sum_i |x_i|^p, \right\} x_i \ge 0. \tag{2.10}$$

Equation (2.10) is difficult at best to solve directly. Rather, one may express this as a constrained minimization by using known (or estimated) noise statistics (σ_η^2) to specify a confidence region about our solution. An upper bound, consistent with our uncertainty due to noise, is set on the first term of (2.10) and then $\underline{x}$ is adjusted to minimize the second term. The resulting constrained optimization problem is easily shown to be

$$\min_{\underline{x}} \sum_i |x_i|^p$$

such that

$$||\underline{b} - \boldsymbol{H}\underline{x}||^2 \le \epsilon, x_i \ge 0, 0 < p < 1. \tag{2.11}$$

With σ_η^2 known, we may specify ϵ to give a fixed probability, α, that the true $\boldsymbol{H}\underline{x}$ lies within a distance ϵ from $\underline{b}$, i.e., select ϵ such that

$$P[||\underline{b} - \boldsymbol{H}\underline{x}||^2 < \epsilon \ |\underline{b}] = \alpha \tag{2.12}$$

$\hat{\underline{x}}$ is then the most sparse vector that maps into this confidence region given by the hypersphere of radius ϵ.

Though (2.11) involves a quadratic constraint rather than the linear inequality constraints of (1.3), clearly use of $g(\underline{x}) = \sum_{i=1}^N (x_i)^p$ as the optimization criterion of choice is indicated by MAP estimation with gpG distributed star images corrupted by blur and Gaussian noise. The results presented above provide justification for an l_p quasi-norm approach to star deblurring and demonstrate a duality between deterministic sparse optimization and optimal parameter estimation for distributions which are likely to produce sparse $\underline{x}$.

III. Algorithms for Sparse Star Deblurring

Though the star deblurring problem has been expressed in a conventional constrained nonlinear optimization form, (1.3) presents a difficult problem from an algorithmic point of view. Since $g(\underline{x})$ is not convex, simple gradient search techniques will not locate a global minimum. We believe, in fact, that sparse optimization belongs to the class of "N-P hard" problems, which cannot in general be solved exactly in polynomial time (i.e., in a number of iterations that can be expressed as a polynomial of order N for an N-dimensional problem.) Thus algorithm design must be directed toward methods with faster average performance, or which find good approximations of the exact global solution. We will discuss three methods which take this approach and have proven effective in various applications. Both the l_p Simplex Search and simulated annealing methods have proven effective for star image deblurring. The collapsing polytopes method is discussed in Section III-C as a point of reference as to why existing global concave minimization algorithms are ineffective for high dimensionality problems like image restoration.

A. The l_p Simplex Search Algorithm

In previous work we have shown that solving (1.3) is equivalent to solving the following problem involving linear equality constraints [11], [20].

$$\min_{\boldsymbol{x}} g(\boldsymbol{x}) = \sum_{i=1}^N (\boldsymbol{x}_i)^p$$

such that

$$\mathcal{H}\boldsymbol{x} = \boldsymbol{b}, 0 < p < 1, \boldsymbol{x}_i \ge 0 \tag{3.1}$$

where

$$\mathcal{H} = \begin{bmatrix} \boldsymbol{H} & \boldsymbol{I} & 0 \\ \boldsymbol{H} & 0 & -\boldsymbol{I} \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} \underline{x} \\ \underline{s}^+ \\ \underline{s}^- \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} \underline{b} + \underline{\epsilon} \\ \underline{b} - \underline{\epsilon} \end{bmatrix}.$$

$\underline{x} \in R^N, \underline{s}^+, \underline{s}^- \in R^M, \underline{b} \in R^M, \mathcal{H}: (2M) \times (N + 2M)$.

I is the $M \times M$ identity matrix, $\underline{s}^+$ and $\underline{s}^-$ are, respectively, slack and surplus variables as commonly employed in linear programming [12], but are not included in the computation of $g(\boldsymbol{x})$ or in the final solution for $\underline{x}$.

Since $g(\boldsymbol{x})$ is concave over $\boldsymbol{x}_i \ge 0$, the solution to (3.1) must lie at an extreme point of the constraint set [11], [20]. These extreme points are called *basic feasible solutions,* and are defined as any $\boldsymbol{x}$ which satisfies $\boldsymbol{H}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x}_i \ge 0$, and contains at most M nonzero elements. This observation has tremendous significance since there are only a finite number of basic feasible solutions. Unfortunately, this set is prohibitively large and grows as the factorial of N. The algorithms presented

here and in the following section are essentially schemes for searching the set of *basic feasible solutions* in an efficient manner to find the optimal solution using something less than an exhaustive search. The problem structure suggests that an algorithm similar to the simplex method used for linear programming (LP) may be effective. This is the basis for the algorithm presented here, which is essentially the same as an approach due to Barrodale and Roberts [21], and which we have called the l_p simplex search.

A basic feasible solution may be computed by partitioning $\mathcal{H}$

$$[\boldsymbol{A}|\boldsymbol{D}] = \mathcal{H}, \text{ where } \boldsymbol{A} \in R^{2M \times 2M} \tag{3.2}$$

multiplying the constraint of (3.1) by $\boldsymbol{A}^{-1}$ leads directly to a basic solution $\boldsymbol{x}_B$

$$[\boldsymbol{I}|\boldsymbol{A}^{-1}\boldsymbol{D}]\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}, \boldsymbol{x}_B = [\boldsymbol{A}^{-1}b, 0, \cdots, 0]^T. \tag{3.3}$$

Any choice of $2M$ columns from $\boldsymbol{H}$ for which$\boldsymbol{A}$ is nonsingular is acceptable for a solution $\boldsymbol{x}_B$. The variables $\boldsymbol{x}_i$ associated with columns of $\boldsymbol{A}$ are termed basic variables, and $\boldsymbol{A}$ the basis. As in the standard LP simplex algorithm, program iterations consist of "pivoting" a column from $\boldsymbol{D}$ into the basis, and the appropriate column from $\boldsymbol{A}$ out of the basis into $\boldsymbol{D}$ to compute a new basic solution adjacent to the previous one. At the ith step of successive pivot operations we will designate $\boldsymbol{A}_i$ as the basis for the current partitioning of $\mathcal{H}$, and $\boldsymbol{x}_{B_i}$ the corresponding basic solution. Two solutions, $\boldsymbol{x}_{B_i}$, and $\boldsymbol{x}_{B_j}$, are said to be neighbors, or adjacent solutions, if one pivot is required to move between them (or equivalently if $\boldsymbol{A}_i$ differs from $\boldsymbol{A}_j$ in only one column.) The set of all neighbors of $\boldsymbol{x}_{B_i}$ is denoted by $N(\boldsymbol{x}_{B_i})$. The utility of the simplex algorithm is that we need not recompute $\boldsymbol{A}^{-1}$ explicitly with each pivot but may use a simple algebraic operation, identical to an LP pivot, directly on (3.3) [12].

Locating a sparse solution is accomplished by selecting leaving and entering columns such that $g(\boldsymbol{x})$ is reduced with each pivot. The sequence of iterations produces a series of solutions with monotonically decreasing cost. The algorithm terminates when all surrounding adjacent solutions are of higher cost than the current one. Note that selection of pivots and computation of the cost are handled differently than in LP.

The steps of the l_p simplex search algorithm are as follows:

1) $i = 0$, find any initial basic feasible solution to (3.1) $\boldsymbol{x}_{B_o} = [\boldsymbol{A}^{-1}\boldsymbol{b}, 0, \cdots 0]^T$, (e.g., the LP phase I solution),
2) Compute the cost, $g(\boldsymbol{x}_{B_j})$, for all bounded $\boldsymbol{x}_{B_j} \in N(\boldsymbol{x}_{B_i})$.
3) If no adjacent solution is of equal or lower cost, terminate, optimum found. Otherwise, if any lower-cost solutions exist, select one and pivot to it. Otherwise, perform an anticycling procedure to pivot to an equal cost solution [12].
4) Increment i and set $\boldsymbol{x}_{B_i}$ to the new solution.
5) Repeat 2) through 4) to termination.

This algorithm has been demonstrated in several different applications to produce excellent approximations to the optimally sparse solution, but due to the nonlinearity of the cost function, $g(\underline{x})$, it may terminate before a true global optimum is located. Results do, however, compare favorably with existing algorithms for sparse star field image restoration and such other diverse applications as sparse beamforming array design, seismic deconvolution, and neuromagnetic image reconstruction [11], [20].

B. Simulated Annealing for Star Deblurring

The l_p Simplex search algorithm described above is very efficient; however, when global optimality is of paramount concern, other more computationally intensive methods are required. Simulated annealing has been shown to be successful for this class of problem, but requires significant computer resources. This section presents the detail of applying simulated annealing to the star reconstruction problem. Another global optimization algorithm which we have investigated, the collapsing polytopes method, is discussed in the following section.

Stochastic relaxation, or simulated annealing as it is often called, has been used by a number of authors in recent years for a wide range of nonlinear combinatorial optimization problems of high dimension [22]–[24]. The l_p simplex search is readily adapted to a simulated annealing algorithm. Here the choice of which adjacent solution to pivot to (step 3 in the l_p Simplex Search) is not always based on strict reduction of the cost, $g(\boldsymbol{x})$. A degree of randomness is introduced in the choice so that less sparse solutions may occasionally be chosen in a given iteration. The degree of randomness is gradually reduced according to an "annealing schedule" until the algorithm stabilizes at the globally optimum solution. Selection of an appropriate annealing schedule is a tradeoff between the required computation time to reach termination, and the degree of solution sparseness.

The Metropolis algorithm [23], [24] as used for a simulated annealing starfield restoration is described below. The following notation is used: $K_i = |N(\boldsymbol{x}_{B_i})|$ is the cardinality of $N(\boldsymbol{x}_{B_i})$, which is a count of the number of neighbors to $\boldsymbol{x}_{B_i}$. $K_{\max} = \max K_i$.

Simulated Annealing l_p Simplex Algorithm Description:

1) $i = 0$, find any initial basic feasible solution to (3.1) $\boldsymbol{x}_{B_o} = [\boldsymbol{A}^{-1}\boldsymbol{b}, 0, \cdots 0]^T$.
2) Select a candidate solution, $\boldsymbol{x}_{b_j}$, to pivot to according the the probability:

$$Pr(\boldsymbol{x}_{B_j}) = \begin{cases} 1/K_{\max}, & \text{if } \boldsymbol{x}_{B_j} \in N(\boldsymbol{x}_{B_i}) \\ 1 - K_i/K_{\max}, & \text{if } \boldsymbol{x}_{B_j} = \boldsymbol{x}_{B_i}. \end{cases}$$

3) $\Delta g = g(\boldsymbol{x}_{B_j}) - g(\boldsymbol{x}_{B_i})$,with $g(.)$ as described in (1.3)
4) If $\Delta g \leq 0$, then $\boldsymbol{x}_{B_i+1} = \boldsymbol{x}_{B_j}$go to step 8).
5) Generate an independent uniformly distributed random variable $r \sim \boldsymbol{U}(0,1)$.
6) If $r \leq \exp\left(\dfrac{-\Delta g}{T_i}\right)$, then $\boldsymbol{x}_{B_i+1} = \boldsymbol{x}_{B_j}$, go to step 8).
7) $\boldsymbol{x}_{B_i+1} = \boldsymbol{x}_{B_i}$.
8) $i = i + 1$, If $i >$ max. iterations, terminate, otherwise update T_i according to the annealing schedule.
9) Go to 2).

This algorithm produces a strongly ergodic Markov chain and converges to the global optimum in the limit, provided the annealing schedule is of the form [24]

$$T_i = \frac{\gamma}{\log(i + n_0 + 1),} \qquad i = 0, 1, 2, \cdots \tag{3.4}$$

where n_0 is any parameter $1 \leq n_0 \leq \infty$, and γ is a constant which, though noncritical, is generally chosen to be proportional to the number of states in the system.

Since the simulated annealing algorithm exhibits only asymptotic convergence, and yet the set of basic feasible solutions is finite, this algorithm is of practical importance only if its finite time behavior yields improved solutions over the deterministic simplex search. Mitra *et al.* analyzed the finite time behavior of the algorithm, and computed a bound for the deviation between the optimal cost and the finite time cost [24]. For a finite sequence, the terminal result is not guaranteed to be the lowest cost solution in the sequence due to the random search. Therefore, truncated procedures can improve performance by maintaining memory storage of the lowest cost solution achieved up to the current iteration. This storage is updated only when the current result is better than all previous points in the sequence. Our experiments verify that the truncated sequence produces improved results over the deterministic l_p simplex search.

C. Collapsing Polytopes Method

Equation (3.1) describes a system which belongs to the class of linearly constrained concave minimization problems which have received significant attention in the literature of operations research. Many related algorithms have been proposed which produce exact global concave minimizations using exterior point search techniques, and which are on average more efficient than an exhaustive search. Since these appear to be the only known finite, efficient methods for exact solutions they should clearly be investigated. We have implemented and evaluated one of the more promising methods [25] representative of this group for its applicability to star deblurring.

The Falk and Hoffman collapsing polytopes algorithm is a finite procedure which solves a sequence of related linear programming problems by pivoting operations [25]. It is guaranteed to terminate at the true global solution and incorporates a procedure which efficiently eliminates the need to search much of the feasible region. The algorithm is often significantly more efficient than an exhaustive search, though this is not guaranteed, and cases can be contrived for which no improvement is realized.

The approach of this algorithm is to extend the dimensionality of (3.1) by augmenting $\boldsymbol{x}$ with a single additional variable, y. The feasible solution set for the augmented problem is in R^{2N+1} space, while the feasible set for the original problem (3.1) is the $2N$ dimensional hyperplane in this space described by $y = 0$. A polytope in R^{2N+1} is formed which encloses the original solution space in one of its faces, which is in the plane $y = 0$. This polytope is iteratively collapsed by a geometric procedure to more closely bound the original solution space until its lowest cost vertex in the plane $y = 0$ corresponds to a basic solution of the original unaugmented problem. Since the polytope face on $y = 0$ is a convex set, if its lowest cost vertex is a basic solution to (3.1), it must be the global solution for the concave cost function $g(\boldsymbol{x})$.

$g(\boldsymbol{x})$ as defined above is not usable in the collapsing polytopes algorithm because of exterior point evaluations where $x_i < 0$. The l_p quasi-norm is not concave over both positive and negative values of x. This problem was overcome by introducing a new objective function, $c(\boldsymbol{x})$, to approximate $g(\boldsymbol{x})$ which is concave over the entire space, and still provides a good measure of sparseness over nonnegative $\boldsymbol{x}$:

$$c(\boldsymbol{x}) = \sum_i \begin{cases} \dfrac{\exp(\boldsymbol{x}_i) - \exp(-\alpha \boldsymbol{x}_i)}{\exp(\boldsymbol{x}_i) + \exp(-\alpha \boldsymbol{x}_i)}, & \boldsymbol{x}_i > 0 \\ \dfrac{(1+\alpha)\boldsymbol{x}_i}{2}, & \boldsymbol{x}_i \leq 0. \end{cases} \tag{3.6}$$

For $\alpha = 1$ this function is recognized as the hyperbolic tangent. For $\boldsymbol{x} > 0$ and as α becomes large, this function approaches the indicator function which, as discussed above, is an obvious measure of sparseness.

To evaluate the algorithm it shall be compared with the l_pSimplex Search. This comparison is, however, difficult and can only provide insight rather than an exact ordering since the convergence criteria are different for each algorithm and one is a global method, while the other provides approximately optimum solutions. Also, no exact analysis is available on the expected number of iterations for either algorithm. Both approaches, however, use linear system extreme point combinatorial searches by pivoting operations.

The collapsing polytopes algorithm, as compared with the l_pSimplex Search, was found to be prohibitively complex and costly for star deblurring for the following reasons:

1) The order of the system to be solved is significantly increased. Positivity constraints must be explicitly included in the system equations. The tableau created for the l_p Simplex Search is of dimension $(2M) \times (N + 2M + 1)$ (see (3.1)), while for collapsing polytopes the system is $(2M + N) \times (2N + 2M + 1)$. Computations per pivot are increased by a factor of approximately N^2, while the number of basic feasible solutions, and potentially the number of iterations, is increased at a factorial rate.
2) Computational overhead per iteration is significantly increased. The search procedure requires maintenance of two sorted lists of all basic solutions previously evaluated. Each iteration involves pivoting to an adjacent solution, a list search, creation of list entries, and traversal of a tree structure by solving a new nested linear sub program of order $(2M+N) \times (2N+2M+1)$. The l_p Simplex Search on the other hand requires only a single pivot per iteration.
3) Memory usage is significantly increased. Both the increased problem dimensionality and the lists add to memory requirements. Each list entry requires $2(N + 2M) + 4$ bytes of storage, and at least one entry is added per iteration. List growth alone can prohibit large problems encountered in star deblurring.

In our experiments the collapsing polytopes methods was successful only for very small problems, relative to the size of star deblurring examples shown in Section V. Collapsing polytopes appears typical of other concave minimization algorithms in its complexity and computational demands. Based on these observations, we conclude that though they are theoretically significant, available linear constrained concave minimization algorithms are impractical for image processing. While this algorithm is not suited for the application of star deconvolution, its initial solution to the center of the polytope often gave a very good starting approximation to the maximally sparse solution. This part of the algorithm could be applied to the l_p Simplex Search for a good initial "guess."

IV. Comparison with the Clean Algorithm

Perhaps the most widely accepted algorithm for incoherent star field deblurring is CLEAN, introduced in 1974 by Högbom [6]. It has been used successfully for deconvolving atmospheric blur and imaging system point spread functions in optical telescopy, and has been widely used in radio telescopic image reconstruction. A CLEAN iteration consists essentially of locating the peak blurred image intensity value, then subtracting a scaled copy of the known point spread function, centered on the peak, from the image. This process is repeated until the peak in the residual image is below a predetermined error tolerance limit. This process has also been called "iterative beam subtraction" when the point spread function corresponds to the beam response pattern of a radio telescopic antenna array. The "cleaned" image consists of nonzero values only in the locations corresponding the peaks which were processed [6], [26].

This procedure has similarities with the l_p simplex search. Both methods are "designed" to operate on sparse sources, that is when the true image consists largely of blank sky [6], [26]. In fact, if this model is inappropriate, neither method may yield acceptable results. Since CLEAN iterations terminate when restored image error drops below a specified limit, it may also be cast in the form of the constrained optimization of (1.2). CLEAN, however, was proposed as an *ad hoc* procedure, with no optimization theoretic basis. Although it clearly sharpens processed images, it is unclear what underlying implicit objective function, $g(\underline{x})$, is active.

It has more recently been shown that the underlying objective in CLEAN approximates l_1 minimization [26]. A "cleaned" image minimizes the sum of pixel intensities within the error constraints

$$\min_{\underline{x}} \sum_{i=1}^{N} x_i$$

such that

$$|\boldsymbol{H}\underline{x} - \underline{b}| \leq \underline{\epsilon}, x_i \geq 0. \tag{4.1}$$

This is readily seen as a special case of the l_p optimization provided by the l_p simplex search. For example, Stark points out the enhanced resolution and generally sparse results in images restored by solving a system equivalent to (4.1) [2]. l_1 minimization however cannot in general achieve the *maximally* sparse result. We have shown that solutions continue to be more sparse as p decreases from 1 to some p_0, at which time (2.1) yields the maximally sparse result [11]. Thus CLEAN, having been designed to recover a true image consisting a few stars in a blank sky, can fall short of its stated objective. In the following section we present an example where the l_p simplex search outperforms CLEAN and restores resolution with fewer extraneous artifacts.

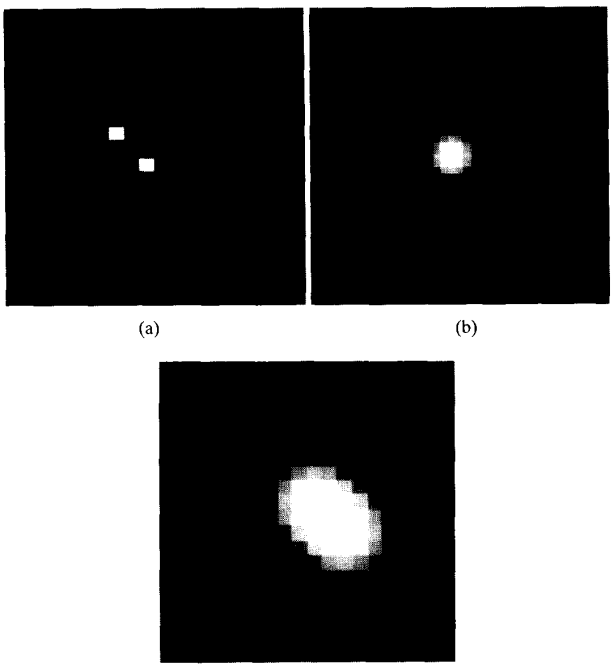
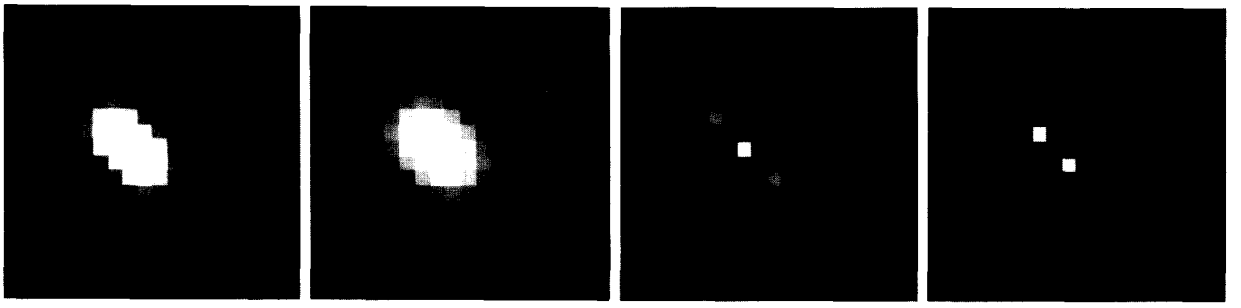


Fig. 3. Synthetic star pair image of 20x20 pixels (a) unblurred image. (b) Point spread function (c) Blurred and noisy image used for processing.

V. Results

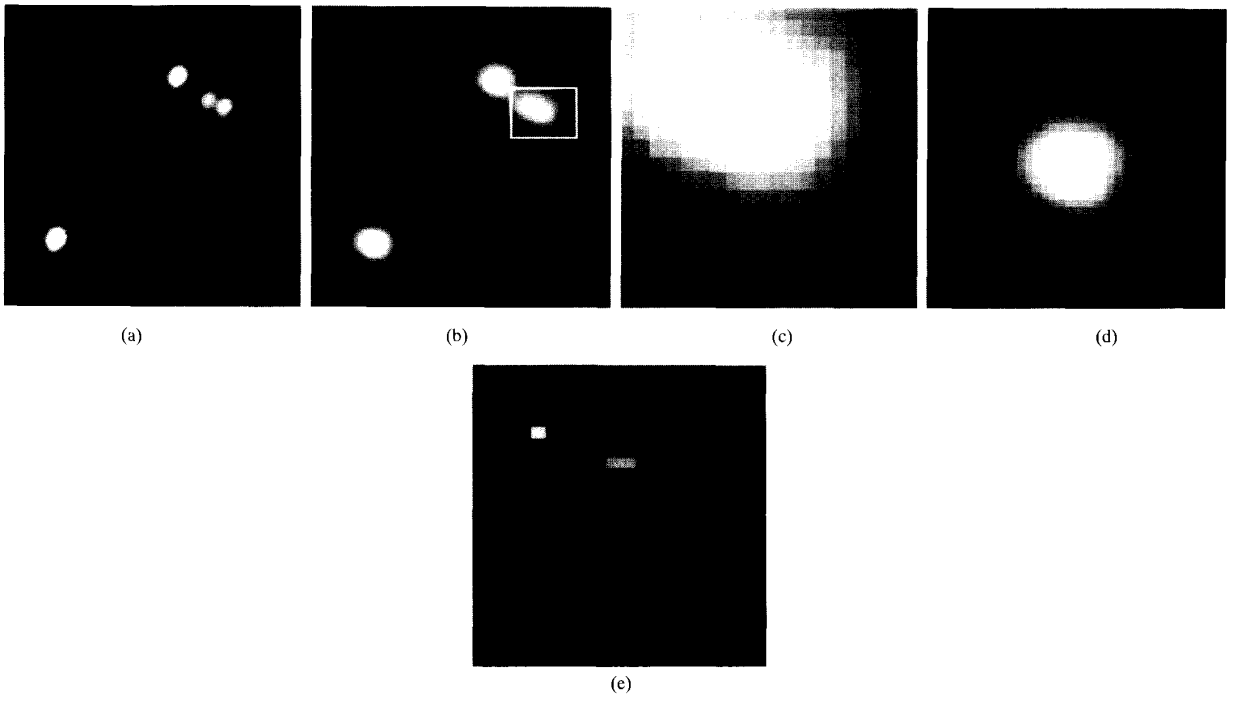
In this section we present experimental results of star field image restoration using the l_p simplex search. Figs. 3 and 4 are from a synthetic image case and give a performance comparison between the l_p simplex search, algebraic reconstruction technique (ART), maximum entropy, and CLEAN reconstructions. Figs. 5 and 6 demonstrate sparse reconstructions from actual telescopic star images.

Fig. 3(a) shows the original ideal image of a synthetic star pair in a black background. Both stars have equal intensity of 1.0 and are located at pixels (9, 9) and (11, 11) respectively in the 20×20 pixel image. The 39×39 pixel point spread shown in Fig. 3(b), a 2-D Gaussian function with a standard deviation of 1.5 pixels, is used to blur the image. Fig. 3(c) presents the final corrupted image truncated to 20×20 pixels with the blurred star pair and additive Chi-squared distributed noise (one degree of freedom). The noise is statistically independent pixel to pixel and was generated by squaring Gaussian samples with a standard deviation of .0125.



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

Fig. 4. Reconstructions of Figure 4(c). using three algorithms: (a) Minimum norm using the ART with relaxation parameter $\lambda = 0.2$ and mean square error residual at convergence of 0.00353. (b) Maximum entropy using multiplicative ART with relaxation parameter $\lambda = 0.2$ and mean square error residual at convergence of .00345. (c) CLEAN with error threshold $\epsilon = 0.02$. (d) l_p Simplex with $p = 1/8$ and error threshold $\epsilon = 0.02$. Note the correct resolution and placement of stars in (d).



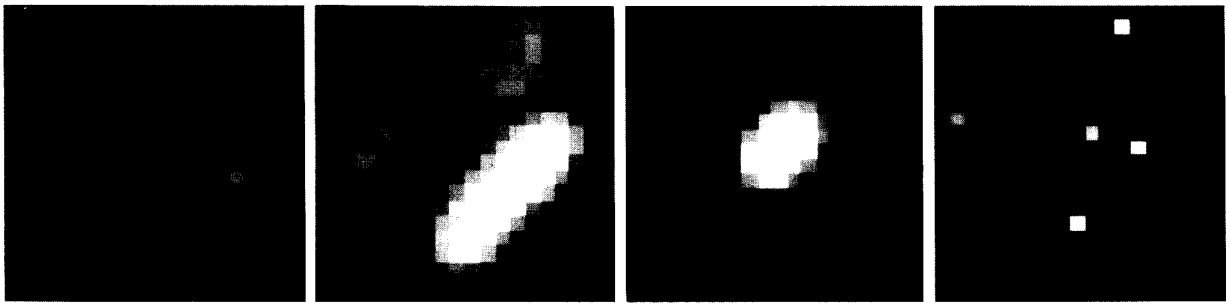
(a) (b) (c) (d)

(e)

Fig. 5. Reconstruction of a binary star near globular cluster M67. Resolution is 1.5 arcseconds/pixel. (a) Resolved image under good seeing conditions. (b) Blurred image. (c) Expanded view of the star pair used in processing. (d) Estimated point spread function from a nearby star. (e) l_p Simplex restoration of the processed region.

The results of various reconstruction methods as applied to Fig. 3(c) are presented in Fig. 4. A minimum norm reconstruction using the ART algorithm [3] is shown in Fig. 4(a). This algorithm produces the minimum energy image consistent with the observed data. The algorithm incorrectly adds structure (concentric rings, etc.) which is not present in the original image. The maximum entropy result shown in Fig. 4(b) was generated with a multiplicative ART algorithm [3], and clearly demonstrates the "maximally smooth" characteristics of this type of reconstruction. Note that the individual stars are not resolved. Fig. 4(c) shows that CLEAN resolves isolated stars but incorrectly places a bright star midway between the true locations, and then two stars more widely separated to reduce the residual error. This behavior is typical of CLEAN whenever blurring is sufficient to smear objects into a single intensity peak. Fig. 4(d) demonstrates the superiority of the l_p Simplex method which resolved the two stars perfectly. The absolute pixel error threshold, ϵ, used in both the CLEAN and l_p Simplex reconstructions was 0.02.

Fig. 5 demonstrates reconstruction of a blurred binary star from an actual telescope image. Fig. 5(a) and 5(b) shows a star grouping near globular cluster M67 under excellent and poor viewing conditions, respectively. Note how the upper right binary star is resolvable in Fig. 5(a), while in 5(b) it is blurred by atmospheric turbulence to the point of exhibiting a single intensity peak. The "boxed" region of Fig. 5(b) indicates the



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

Fig. 6. Reconstruction of a star group seen in selected area 110 as cited in [24]. Image is from CWRU/NOAO observatory with a 36/24 Burrel Schmidt telescope having 2.1 arcseconds/pixel resolution. (a) Star group under more favorable viewing conditions. (b) 20×20 blurred image which is processed. (c) estimated point spread function from a nearby star. (d) Reconstructed image using l_p Simplex search with $p = 1/8$. and error threshold $\epsilon = 0.3$.

portion of this image which was processed for reconstruction, and is shown in expanded view in Fig. 5(c). The point spread function estimate (Fig. 5(d)) used in deconvolving the image is a neighboring isolated star. Fig. 5(e) shows the result of the l_p Simplex algorithm. An interesting feature is the two adjacent pixels which represent the right star in the binary. We speculate that the star's true position was a point between the two pixels, thus the algorithm smeared the energy between the two bins. In this example, the sharper optical image of Fig. 5(a) serves as a "ground truth" frame of reference to confirm that the reconstruction of the blurred region is in fact correct. This may lend confidence in applying the method to blurred images where the true underlying star field is unknown.

Reconstruction of a second example star group (see Landolt [27] for sky coordinates) is presented in Fig. 6. This case includes blurring from atmospheric turbulence, motion due to improper telescope tracking, and misfocus. The first image (Fig. 6(a)) was taken with a shorter exposure time under somewhat better conditions and shows several resolved stars as well as an elongated cluster in the lower right which contains an indeterminate number of stars. Fig. 6(b) contains the image which was processed and exhibits blurring of some of the stars which are resolved in Fig. 6(a). The point spread function taken from a nearby isolated star (Fig. 6(c)) has an interesting structure like an elongated "donut" due to misfocus an tracking error. Note that this structure can also be seen in the blurred image. In Fig. 6(d) we see the results of the l_p Simplex method. It places stars where we expected for the blurred isolated stars and it determined that there are two main stars in the elongated cluster as well as a third dimmer star placed slightly above them.

VI. Conclusion

The above examples demonstrate that maximally sparse restoration is a promising approach for star deblurring. It is ideally suited to cases where only point sources, and no distributed objects, exist in the true image field. In any restoration problem, the form of a restored image is highly dependent on the objective function (explicit or implicit) underlying the algorithm which is used. The commonly-invoked criteria (least squares, maximum entropy, etc.) impose a smoothing operation on the solution image. One may argue that when a sparse source image model is used, the only justifiable objective is a measure of sparseness. The l_p Simplex Search is a practical maximally-sparse algorithm, and outperforms CLEAN, ART, and maximum entropy methods in cases of severe blurring. Though this algorithm yields good approximations of the maximally sparse solution which are acceptable for star image resolution enhancement, it does not assure location of the global optimum. For cases which require a global optimum, we have shown how simulated annealing can be applied to the simplex search. Though this is computationally intensive, it does produce superior results. The collapsing polytopes method was evaluated as typical of published concave minimization algorithms which seem applicable to this problem, but was found to be extremely impractical for high dimensionality problems found in image restoration.

Acknowledgment

The authors would like to acknowledge the invaluable assistance of Dr. J. Ward Moody of the Brigham Young University Department of Physics and Astronomy, who advised them on the astronomical aspects of the project and provided the star image data used in the examples. Also, Matthew Collier, Timothy Jeffcoat, John Dougall, and Douglas Elsmore have contributed to this paper in various capacities as research assistants.

References

[1] R. Chellappa and A. A. Sawchuk, Eds., *Digital Image Processing and Analysis, Digital Image Processing*, vol. I, Los Alamitos, CA: IEEE Computer Society Press, 1985, pp. 557–695.

[2] H. Stark, *Image Recovery: Theory and Application.* New York: Academic, 1987.

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

[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.*, vol. 191, pp. 69–79, 1980.

[5] S. F. Gull and J. Skilling, "Maximum entropy method in image processing," *Proc. Inst. Elect. Eng.*, vol. 131, pt. F, no. 6, pp. 646–659, 1984.

[6] J. A. Högbom, "Aperature synthesis with a non-regular distribution of interferometer baselines," *Astron. Astrophys. Suppl.*, vol. 15, p. 417, 1974.

[7] L. Sica, "Use of a sharpness criterion in correcting images degraded by random distortion," *J. Opt. Soc. Amer. A*, vol. 5, no. 9, pp. 1492–1501, 1988.
[8] C. K. Rushforth and R. L. Frost, "Comparison of some algorithms for reconstructing space-limited images," *J. Opt. Soc. Amer.*, vol. 70, no. 12, pp. 1539–1544.
[9] I. R. King, "The profile of a star image," *Pub. Astron. Soc. Pacific*, vol. 86, Apr. 1971.
[10] B. Jeffs, R. Leahy, and M. Singh, "An evaluation of methods for neuromagnetic image reconstruction,"*IEEE Trans. Biomed. Eng.*, vol. BME-34, pp. 713–723, 1987.
[11] B. D. Jeffs, "Maximally sparse constrained optimization for signal processing application," Ph.D. dissertation, Univ. of Southern California, Jan. 1989.
[12] D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed., Reading, Mass. Addison-Wesley, 1984.
[13] W. Gray, "Variable norm deconvolution," Ph.D. Thesis, Stanford Univ., 1979.
[14] M. T. Subbotin, "On the law of frequency of errors," *Mat. Sbornik*, vol. 31, no. 1, pp. 296–301, 1923.
[15] H. Miller and J. B. Thomas, "Detector for discrete-time signals in non-gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 241–250, Mar. 1972.
[16] J. B. McDonald, "Partially adaptive estimation of ARMA time series models," *Int. J. Forecasting*, no. 5, pp. 217–230, 1989.
[17] R. Wiggins, "Entropy guided deconvolution," *Geophys.*, vol. 50, pp. 2720–2726, 1985.
[18] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis, II.* New York: Academic, 1981.
[19] T. T. Pham and R. J. P. deFigueiredo, "Maximum likelihood estimation of a class of non-gaussian densities with application to l_p deconvolution," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 37, pp. 73–82, Jan. 1989.
[20] R. Leahy and B. Jeffs, "On the design of maximally sparse beamforming arrays," *IEEE Trans. Anten. and Propagation*, vol. 39, pp. 1178–1187, Aug. 1991.
[21] I. Barrodale and F. Roberts, "Application of mathematical programming to l_p approximation," *Nonlinear Programming*, J. Rosen *et al*, Eds. New York: Academic, 1970.
[22] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
[23] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. PAMI-6, pp. 721–741, 1984.
[24] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, "Convergence and finite-time behaviour of simulated annealing," *Adv. Appl. Prob.*, vol. 18, pp. 747–771, 1986.
[25] J. E. Falk and K. L. Hoffman, "Concave minimization via collapsing polytopes," *Oper. Res.*, vol. 34, no. 6, pp. 919–929, Nov.—Dec. 1986.
[26] K. A. Marsh and J. M. Richardson, "The objective function implicit in the clean algorithm," *Astron. Astrophys.*, vol. 182, pp. 17–178, 1987.
[27] A. Landolt, "UBV photoelectric sequences in the celestial equatorial selected areas 92-115*," *Astron. J.*, vol. 78, p. 959, 1973.

Photos and biographies are not available at time of publication.