# Statistical Imaging in Radio Astronomy via an Expectation-Maximization Algorithm for Structured Covariance Estimation

**Aaron D. Lanterman**

ABSTRACT
Image formation in radio astronomy is often posed as a problem of constructing a nonnegative function from sparse samples of its Fourier transform. We explore an alternative approach which reformulates the problem in terms of estimating the entries of a diagonal covariance matrix from Gaussian data. Maximum-likelihood estimates of the covariance cannot be readily computed analytically; hence we investigate an iterative algorithm originally proposed by Snyder, O'Sullivan, and Miller in the context of radar imaging.
The resulting maximum-likelihood estimates tend to be unacceptably rough due to the ill-posed nature of maximum-likelihood estimation of functions from limited data, so some kind of regularization is needed. We explore penalized likelihoods based on entropy functionals, a roughness penalty proposed by Silverman, and an information-theoretic formulation of Good's roughness penalty crafted by O'Sullivan. We also investigate algorithm variations that perform a generic smoothing step at each iteration.

## 1 Introduction

The insatiable thirst of astronomers for increasingly high resolution images of the cosmos has led to ingenious developments in both instrumentation and the processes employed to extract information from the available raw data. In optical astronomy, system resolution is, at first glance, limited by atmospheric turbulence. Undaunted, researchers have tackled the challenges with a variety of tactics, including flexible mirrors that can rapidly bend to compensate for the fluctuating atmosphere, and sophisticated signal processing algorithms for combining multiple short exposures into a single image. At the comparatively low frequencies of interest in radio astronomy, the resolution of a single radio telescope is significantly limited by the difficulty and extraordinary cost involved in constructing antennas

with large dishes. In the 1940's, this hurdle was overcome by correlating the signals from multiple antennas to form an interferometer. Diverse data can be collected by mounting the antennas on tracks or exploiting the rotation of the earth to take interferometric measurements at different positions [RH60, Fom73].

Since the genesis of interferometric radio astronomy, researchers have exploited the connection between the correlation measurements and the Fourier transform of the object being imaged. Basically, each correlation measurement gives an estimate of one point of that Fourier transform. The position of the point in Fourier space depends on the vector separation of the two antennas associated with that measurement. For instance, an array of 27 elements (unequally spaced to avoid redundant position differences) provides $27^2 = 729$ Fourier points. By changing the position of the array elements (either by physically moving the elements or by waiting for the Earth to rotate), additional points may be obtained, but it is almost impossible to obtain all the points needed to form a good image by direct Fourier inversion. Simply setting the missing points to zero and applying an inverse Fourier transform yields a blurry image with complicated sidelobe structures, often called the "dirty map." In addition, negative pixel values are often obtained, even though the intensity should be nonnegative. Radio astronomers rely on two main tools, the CLEAN algorithm [H74, Sch78] and maximum entropy [Abl74, Pon80], to produce useful "clean maps." Although the underlying mapping between the correlation data and the intensity map is linear, CLEAN and maximum entropy algorithms are both nonlinear forms of processing.

This chapter explores a radically different approach based on statistical estimation theory. It does not require or explicitly exploit the Fourier relationships which have entirely dominated radio interferometry for the past fifty years. The method can deal with any reasonable linear relationship between the astronomical source and the raw measurements; the fact that these relationships are complex exponentials is somewhat incidental. Under the statistical model, the pixels that form the sky emit white Gaussian processes. The antenna elements see a mixture of these processes through a linear transformation and additive receiver noise. The signals seen at the antenna elements are then another set of Gaussian process, with correlations between the components. One can then simply write down the likelihood of the data given the unknown powers and maximize the likelihood with respect to the unknown parameters. This turns out to be a challenging maximization which does not yield to a frontal assault; hence, we adopt an expectation-maximization (EM) algorithm originally proposed by Snyder, O'Sullivan, and Miller [SOM89] in the different but related context of imaging diffuse radar targets.

The CLEAN algorithm, and traditional maximum-entropy algorithms, require the investigator to choose a variety of parameters; the results of these algorithms are often highly dependent on the skill of the practitioner

in choosing these parameters. By contrast, in adopting the maximum-likelihood framework, one can "turn the crank" of the EM algorithm without having to tune any such operational parameters. Our new approach also fully exploits all statistical knowledge about the data collection.

One of the advantages of maximum-entropy techniques in traditional radio astronomy formulations is that the entropy functional ensures nonnegative estimates. The maximum-likelihood approach has this same advantage; nonnegativity is automatically guaranteed by the form of the EM iterations.

To our knowledge, Leshem and van der Veen [LvdV00] offer the only other reported work on a maximum-likelihood approach to radio astronomy. They consider imaging in the presence of strong man-made radio interference, for instance, from communication satellites; hence, their primary interest lies in performing maximum-likelihood inference for specific point sources. This yields close ties to the direction-of-arrival literature [JD93]. They present a simplifying approximation to the loglikelihood and a coordinate descent algorithm which treats already-discovered point sources as colored noise. This differs, in both goal and execution, from our EM algorithm for pixel-based imaging. Thus, this chapter and [LvdV00] complement one another.

Section 2 presents our statistical formulation of the radio astronomy problem, and Section 3 presents the EM algorithm for computing estimates using this model, as well as some simulation results. These experiments illustrate the need for regularization; various possibilities are explored in Section 4. Section 5 suggests some avenues for future exploration.

The literature on radio astronomy is vast and rich. We have found the collections of papers edited by Goldsmith [Gol88] and Felli and Spencer [FS89], as well as the text by Thompson, Moran, and Swenson [TMS88], to be particularly helpful.

## 2 Problem Formulation

Suppose the sensor array has $K$ elements. Since radio astronomy antennas are quite expensive, it is common to either mount the antennas on tracks which allow them to be moved or to simply let the earth rotate, which places them at a new effective look position. Suppose we take data at $M$ such look positions and that we collect $N$ snapshots at each position.

Anticipating discretization for computer implementation, we suppose that the sky is made up of $I$ pixels. We will use a single index $i = 1, \ldots, I$ to run over the the entire two-dimensional pixel array. Let $\phi_i$ and $\theta_i$ denote the angles that pixel $i$ form with respect to a line perpendicular to and crossing the center of the array. Assume each pixel $i$ radiates a complex, white Gaussian process $c_i(n, m), n = 1, \ldots, N, m = 1, \ldots, M$ with

unknown variance $\sigma_i^2$. Let $\boldsymbol{\sigma}^2 = [\sigma_1^2, \ldots, \sigma_I^2]^T$ and $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma}^2)$.

Assuming that $\phi_i$ and $\theta_i$ are small so that small-angle approximations to trigonometric functions apply, each sensor sees the signal from $c_i$ with an associated phase shift $\exp\{j2\pi[x_k(m)\phi_i + y_k(m)\theta_i]\}$, where $(x_k(m), y_k(m))$ are the coordinates of sensor $k$ at look position $m$. We can write the data received at sensor $k$

$$r_k(n,m) = \sum_{i=1}^{I} c_i(n,m) \exp\{j2\pi[x_k(m)\phi_i + y_k(m)\theta_i]\} + w_k(n,m), \quad (1.1)$$

where $w_k(n,m)$ is white, complex, zero-mean Gaussian receiver noise with variance $N_0$.

Let $\mathbf{r} = [r_1, \ldots, r_K]^T$. For convenience, we form matrices of complex exponentials $\boldsymbol{\Gamma}^H(m), m = 1, \ldots, M$, and express (1.1) for all $k$ as

$$\mathbf{r}(n,m) = \boldsymbol{\Gamma}^H(m)\mathbf{c}(n,m) + \mathbf{w}(n,m), \quad (1.2)$$

where $\mathbf{c} = [c_1, \ldots, c_I]^T$ and $\mathbf{w} = [w_1, \ldots, w_K]^T$ are independent and identically distributed as $CN(0, \boldsymbol{\Sigma})$ and $CN(0, N_0\mathbf{I})$, independent with respect to each other, and independent with respect to $n$ and $m$.

Note that for all $n$, $\mathbf{r}(n,m) \sim N(0, \mathbf{K_r}(m))$, where

$$\mathbf{K_r}(m) = \boldsymbol{\Gamma}^H(m)\boldsymbol{\Sigma}\boldsymbol{\Gamma}(m) + N_0\mathbf{I}. \quad (1.3)$$

Estimating the intensity map $\boldsymbol{\Sigma}$ brings us to the structured covariance estimation waters first charted by Burg *et al.* [BLW82]. The loglikelihood for the data is

$$L_{id}(\boldsymbol{\Sigma}) = -N \sum_{m=1}^{M} \ln\det \mathbf{K_r}(m) - \sum_{m=1}^{M}\sum_{n=1}^{N} \mathbf{r}^H(n,m)\mathbf{K_r}^{-1}(m)\mathbf{r}(n,m). \quad (1.4)$$

# 3 An Algorithm for Maximum-Likelihood Imaging

Maximum-likelihood imaging requires us to find the diagonal $\boldsymbol{\Sigma} = diag(\boldsymbol{\sigma}^2)$ that maximizes (1.4). Since no simple solution is evident for this difficult optimization problem, we turn to the expectation-maximization algorithm. The algorithm explored here is a trivial extension of the EM algorithm designed by Snyder, O'Sullivan, and Miller [SOM89] in the context of radar imaging. In the context of the EM algorithm, $\mathbf{r}(n,m)$ is called the *incomplete data* and (1.4) is the *incomplete data loglikelihood*.

To formulate an EM algorithm, one postulates a set of *complete data* which, if available, would make the maximization problem tractable. Here we take the complete data to be $\{\mathbf{c}(n,m), \mathbf{w}(n,m), n = 1, \ldots, N, m =$

$1, \ldots, M\}$, where $\mathbf{c}(n,m) \sim CN(0, \mathbf{\Sigma})$ and $\mathbf{w}(n,m) \sim CN(0, N_0\mathbf{I})$. The classic EM formulation requires that there be a many-to-one mapping from the complete data to the incomplete data; this is provided by (1.2). The complete data loglikelihood is

$$\begin{aligned} L_{cd}(\mathbf{\Sigma}) &= -NM \ln\det \mathbf{\Sigma} - \sum_{m=1}^{M}\sum_{n=1}^{N} \mathbf{c}^H(n,m)\mathbf{\Sigma}^{-1}\mathbf{c}(n,m) \\ &= -NM\sum_{i=1}^{I} \ln \sigma_i^2 - \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{i=1}^{I} \frac{|c_i(n,m)|^2}{\sigma_i^2}. \end{aligned} \tag{1.5}$$

Let $Q$ denote the expectation of the complete data loglikelihood given the incomplete data and the estimate from the previous iteration:

$$\begin{aligned} Q[\mathbf{\Sigma}|\mathbf{\Sigma}^{(old)}] &\stackrel{\text{df}}{=} E[L_{cd}(\mathbf{\Sigma})|\mathbf{\Sigma}^{(old)}, \mathbf{r}] \\ &= -NM\sum_{i=1}^{I} \ln\sigma_i^2 - N\sum_{m=1}^{M}\sum_{i=1}^{I} \frac{E[|c_i(m)|^2|\mathbf{\Sigma}^{(old)}, \mathbf{r}]}{\sigma_i^2}. \end{aligned} \tag{1.6}$$

Notice that in the second term, since the $N$ snapshots of $\mathbf{c}_i$ are independent, we have not notated the explicit dependence of $\mathbf{c}_i$ on $n$ in the expectation.

At each iteration of the EM algorithm, we get the new estimate by maximizing $Q$:

$$\mathbf{\Sigma}^{(new)} = \arg\max_{\mathbf{\Sigma}} Q[\mathbf{\Sigma}|\mathbf{\Sigma}^{(old)}]. \tag{1.7}$$

The derivative of Equation (1.6) with respect to $\sigma_i^2$, set to zero, is

$$-NM\frac{1}{\sigma_i^2} + N\sum_{m=1}^{M} \frac{E[c_i(m)^2|\mathbf{\Sigma}^{(old)}, \mathbf{r}]}{(\sigma_i^2)^2} = 0, \tag{1.8}$$

which yields the simple update

$$\sigma_i^{2(new)} = \frac{1}{M} E[c_i(m)^2|\mathbf{\Sigma}^{(old)}, \mathbf{r}]. \tag{1.9}$$

Computing the expectation in Equation (1.9) is a standard problem addressed in estimation theory texts (see, for instance, Equation 7.112 on p. 303 of Scharf [Sch91] or Equations V.B.21 and 7.B.22 on p. 221 of Poor [Poo94]). Its solution results in the explicit update

$$\begin{aligned} \sigma_i^{2(new)} &= \sigma_i^{2(old)} - \frac{[\sigma_i^{2(old)}]^2}{M}\sum_{m=1}^{M} [\mathbf{\Gamma}(m)\mathbf{K}^{-1}(m)\mathbf{\Gamma}^H(m) \\ &\quad -\mathbf{\Gamma}(m)\mathbf{K}^{-1}(m)\mathbf{S}(m)\mathbf{K}^{-1}(m)\mathbf{\Gamma}^H]_{ii}, \end{aligned} \tag{1.10}$$

where

$$\mathbf{K}(m) = \mathbf{\Gamma}^H(m)\mathbf{\Sigma}^{(old)}\mathbf{\Gamma}(m) + N_0\mathbf{I} \tag{1.11}$$

and

$$\mathbf{S}(m) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{r}(n,m)\mathbf{r}^H(n,m). \tag{1.12}$$

Notice that the data only enters into the inference via its empirical covariance $\mathbf{S}$. We have written (1.10) in a way to show the symmetric structure of the matrix computations. In implementation, it is more efficient to compute the term in brackets by calculating $\Xi(m) = \mathbf{\Gamma}(m)\mathbf{K}^{-1}(m)$ followed by $\Xi(m)[\mathbf{\Gamma}^H(m) - \mathbf{S}(m)\Xi^H(m)]$. Since only the diagonal terms are needed, the final step only requires $I$ inner products.

Taking $N = 1$ and $M = 1$ yields the original algorithm [SOM89] for forming radar images of diffuse targets. There, $\mathbf{c}$ is the reflectance of a radar scatterer, $\boldsymbol{\sigma}^2$ is called its *scattering function*, and $\mathbf{\Gamma}$ contains time-shifted and doppler-shifted versions of the transmitted waveform.

### *3.1 Simulations*

Our simulations are meant to illustrate the overall operation of the EM algorithm, and are not intended to represent any particular real-world scenario. We assume a 27-element array modeled after the Very Large Array in New Mexico [NTE83]. The array consists of three arms arranged in a Y pattern, with equal angular spacing between each arm. Each arm has nine elements. The distance of the $n$th antenna on each arm from the center of the array is proportional to $n^{1.716}$. The real VLA's elements are mounted on tracks, so the overall size of the array can be changed. Here, we assume the array is set at its largest configuration, in which the furthest elements on each arm are 21 km away from the center.

For simplicity, our simulated array departs from the real array in New Mexico in that we suppose the array can be rotated around its center to get different looks (or equivalently, the array is centered at one of the Earth's poles). We will suppose that $M = 5$ different looks are taken, and that the array rotates 10 degrees between looks, for a total rotation of 40 degrees. Figure 1 shows the Fourier sampling pattern this generates in the traditional radio astronomy paradigm. At each look $m$, $N = 500$ snapshots are taken to form the empirical covariance matrix $\mathbf{S}(m)$.

The top row of Figure 2 shows two ideal intensity functions used in our experiments. The bottom row shows the simple estimate

$$\hat{\sigma}_i^2 = \text{real}\left\{\left[\sum_{m=1}^{M}\mathbf{\Gamma}(m)\mathbf{S}(m)\mathbf{\Gamma}^H(m)\right]_{ii}\right\}, \tag{1.13}$$

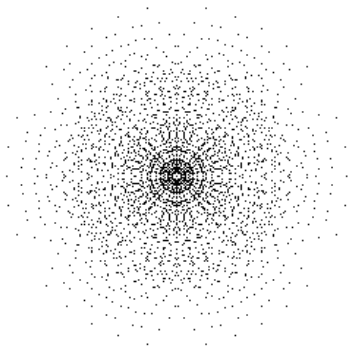

FIGURE 1. Fourier sampling pattern associated with the simulated scenario.

which corresponds to the "dirty map" of traditional radio astronomy.

Figure 3 show the results of EM iterations (1.10) on data simulated from these ideal intensity functions using the described set-up.

# 4 Regularization Techniques

Although the estimates shown in Figure 3 are quite sharp, and the background chaff becomes less evident with increasing iterations, the reconstructed objects suffer from a spiky, noisy appearance. The roughness associated with increasing EM iterations is indicative of the ill-posed nature of maximum-likelihood estimation of functions from limited data. In most numerical algorithms, accuracy improves as the discretization is refined. Here, the opposite is true; as the grid is refined, the problem becomes increasingly ill-posed and the solutions increasingly ill-behaved. This phenomena, called *dimensional instability* by Tapia and Thompson [TT78], would manifest itself in any algorithm that maximizes the loglikelihood.

These issues have been extensively studied in the context of Poisson intensity estimation in applications such as medical imaging (see [SM85], [SMLTP87], and Chapter 3 of [SM91]). Several solutions have been proposed for Poisson imaging which we can adapt to our radio astronomy problem.

## *4.1 The Method of Sieves*

One approach is to restrict the solution to lie in a restricted subset called a *sieve* [Gre81]. One possibility is to require it to be a weighted sum of basis functions. Gaussian basis functions yielded much success in emission tomography [PS88, SMLTP87, PS91]. In the radio astronomy problem

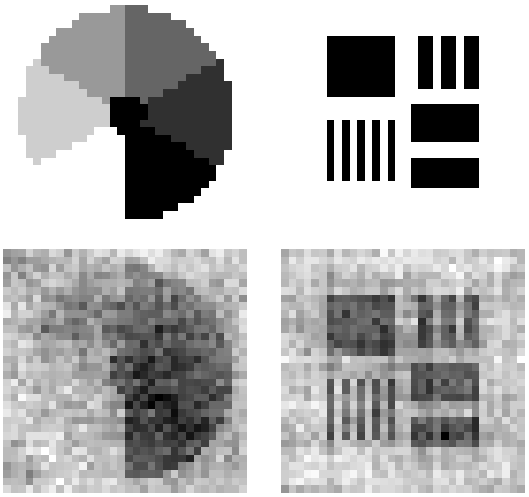

FIGURE 2. Top row: Two ideal intensity functions used in the simulations. Bottom row: Traditional "dirty maps" formed from simulated data.

and the related radar imaging problem originally considered by Snyder, O'Sullivan, and Miller [SOM89], the EM algorithm of Section 3 can be extended to incorporate such a sieve constraint. Moulin and co-workers [MOS92, OSPM92] present this extended algorithm (for N=1, M=1), along with specific examples employing B-splines for radar imaging. Radio astronomers employing the CLEAN algorithm typically smooth the result by convolving it with a so-called "clean beam." This clean beam may provide a natural starting point for choosing the sieve in maximum-likelihood estimation. Wavelet thresholding techniques have also been explored for denoising speckled radar imagery [Mou93].

## *4.2 Regularization via Penalties*

Another approach to regularization is to subtract a penalty $\Phi(\mathbf{\Sigma})$ which discourages unacceptably rough estimates and to maximize the penalized likelihood

$$P(\mathbf{\Sigma}) = L_{id}(\mathbf{\Sigma}) - \alpha\Phi(\mathbf{\Sigma}). \tag{1.14}$$

Bayesians may think of this as maximum-posterior estimation using a (usually improper) prior proportional to $\exp[-\alpha\Phi(\mathbf{\Sigma})]$.

The EM algorithm can be easily extended to maximize this penalized

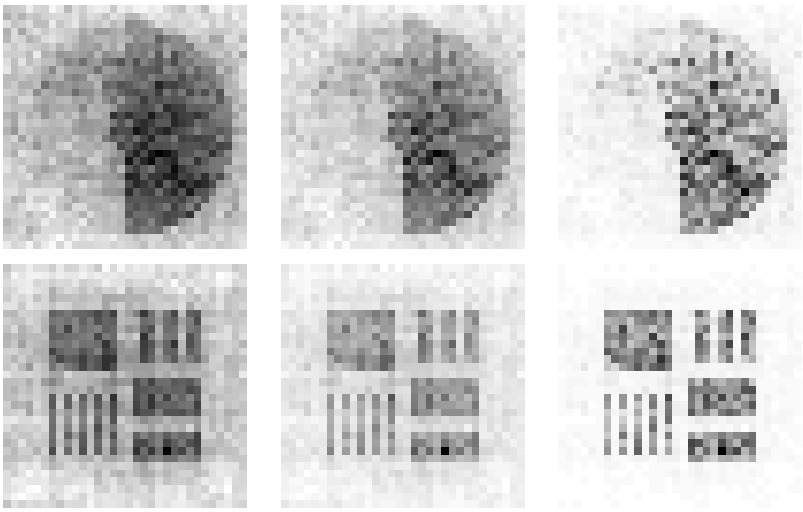

FIGURE 3. Results of the unconstrained EM algorithm for two different data sets (shown in different rows). From left to right, columns show results at 100, 200 and 1000 iterations.

likelihood. The function $Q$ (1.6) is generalized to

$$\begin{aligned} Q_P[\mathbf{\Sigma}|\mathbf{\Sigma}^{(old)}] &\stackrel{\text{df}}{=} E[L_{cd}(\mathbf{\Sigma})|\mathbf{\Sigma}^{(old)}, \mathbf{r}] - \alpha\Phi(\mathbf{\Sigma}) \qquad (1.15) \\ &= -NM\sum_{i=1}^{I} \ln \sigma_i^2 - N\sum_{m=1}^{M}\sum_{i=1}^{I} \frac{E[|c_i(m)|^2|\mathbf{\Sigma}^{(old)}, \mathbf{r}]}{\sigma_i^2} - \alpha\Phi(\mathbf{\Sigma}). \end{aligned}$$

To maximize (1.15), we can take derivatives (analogous to (1.8)) and solve the set of equations:

$$-NM\frac{1}{\sigma_i^2} + N\sum_{m=1}^{M} \frac{E[c_i(m)^2|\mathbf{\Sigma}^{(old)}, \mathbf{r}]}{(\sigma_i^2)^2} - \alpha\frac{\partial\Phi(\mathbf{\Sigma})}{\partial\sigma_i^2} = 0, \qquad (1.16)$$

alternatively written as

$$-\frac{1}{\sigma_i^2} + \frac{\sigma_i^{2(uc)}(\mathbf{\Sigma}^{(old)})}{(\sigma_i^2)^2} - \frac{\alpha}{NM}\frac{\partial\Phi(\mathbf{\Sigma})}{\partial\sigma_i^2} = 0, \qquad (1.17)$$

where $\boldsymbol{\sigma}^{2(uc)}(\mathbf{\Sigma}^{(old)})$ is the result of the *unconstrained* update specified by (1.10). At each iteration, we take the updated value $\mathbf{\Sigma}^{(new)}$ to be the $\mathbf{\Sigma}$ that solves (1.17).

*Entropy Functionals*

Beginning with Frieden [Fri72], numerous authors [ES88, AH91, Mea86, SSB79, SB84, FW78, GD80, WD77] have proposed regularizing a variety of

inverse problems with nonnegativity constraints via the entropy functional

$$\Phi_E(\mathbf{\Sigma}) = \sum_i \sigma_i^2 \ln \sigma_i^2. \tag{1.18}$$

At each iteration, the new $\mathbf{\Sigma}$ is found by finding the zero of

$$-\sigma_i^2 + \sigma_i^{2(uc)}(\mathbf{\Sigma}^{(old)}) - \frac{\alpha}{NM}(\sigma_i^2)^2(1 + \ln \sigma_i^2) \tag{1.19}$$

for each $i$. This is convenient because the solution is decoupled from pixel to pixel. Molina and Ripley [MR89] suggest that entropy "corresponds to a rather peculiar prior, since it depends only on the marginal distribution of grey levels and not on their spatial locations. It is thus surprising that maximum entropy solutions appear smooth in many published examples." Donoho, Johnstone, Hoch, and Stern [DJHS92] offer a highly practical discussion of the how entropy regularization operates in practice. They suggest that nonlinearities of the form induced by (1.19) encourage a "shrinking" of estimates towards a nominal value of $1/e$. Narayan and Nityanada [NN84] study a general class of functions which include (1.18) as a special case and have similar effects on the reconstruction.

It is important to avoid a potential misunderstanding. Maximizing the penalized likelihood (1.14), using the entropy penalty (1.19) and the loglikelihood given by Equation (1.4), is *not* equivalent to "maximum entropy" as traditionally practiced by radio astronomers. In traditional maximum entropy, one maximizes $-\sum_i \sigma_i^2 \ln \sigma_i^2$ subject to agreement with the correlation data. This agreement is most often measured using a simple least-squares criterion; this could be viewed as maximizing a penalized likelihood where the "likelihood" has an implied Gaussian form, with the correlation measurements assumed to be corrupted by additive white Gaussian noise,[1] which is not at all like (1.4).

*Good's Roughness*

Good's roughness penalty [GG71] was originally formulated for smoothing estimates in nonparametric probability density estimation; a thorough analysis in this context is given by Tapia and Thompson [TT78]. Following the suggestion of Snyder and Miller ([SM85], Section II.1), Good's roughness was later applied to closely related problems of Poisson intensity estimation in PET [MR91], SPECT [MM91, BM93, MB93, BMMW94], and optical-sectioning microscopy [JM93].

In the next few subsections, in order to conveniently express the penalties, we will index $\sigma^2$ using two spatial coordinates, as in $\sigma^2_{i_1,i_2}$, $i_1 = 1, \ldots, I_1$, $i_2 = 1, \ldots, I_2$, where $I = I_1 I_2$. The penalties we now explore

[1] The empirical correlation would be more accurately modeled as following a Wishart distribution [GN99].

are most easily interpreted in their original continuous form, so we will also let $\sigma^2(\cdot)$ or $\sigma^2(\cdot,\cdot)$ denote the continuous functions underlying the discrete representations. In one dimension, Good's continuous first-order roughness penalty [GG71] may be written in several equivalent ways:

$$\begin{aligned} 4\int\left[\frac{d}{dx}\sqrt{\sigma^2(x)}\right]^2 dx &= \int \sigma^2(x)\left[\frac{d}{dx}\ln\sigma^2(x)\right]^2 dx \\ = \int \frac{(\sigma^2)'(x)}{\sigma^2(x)}dx &= -\int \sigma^2(x)\frac{d^2}{dx^2}\ln\sigma^2(x)dx. \end{aligned} \tag{1.20}$$

The last equality in (1.20), which was established independently in [O'S95] and [Fri91], follows from integration by parts. Consider the rightmost expression. The penalty is straightforwardly extended to two dimensions (see pp. 155-156 of [SM91] or Section 3 of [MR91]):

$$-\int\int \sigma^2(x,y)\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\ln\sigma^2(x,y)dxdy. \tag{1.21}$$

Discretizing Equation (1.21) yields $\Phi_G(f) =$

$$-\sum_{i_1,i_2}\sigma^2_{i_1,i_2}[\ln\sigma^2_{i_1+1,i_2}+\ln\sigma^2_{i_1-1,i_2}+\ln\sigma^2_{i,i_2+1}+\ln\sigma^2_{i_1,i_2-1}-4\ln\sigma^2_{i_1,i_2}]. \tag{1.22}$$

O'Sullivan [O'S95] noted that the discretized penalty (1.22) has an appealing information-theoretic interpretation in terms of I-divergences between neighboring pixel values.

At each iteration, the new $\mathbf{\Sigma}$ may be found by solving a set of nonlinear difference equations:

$$\begin{aligned} 0 &= -\frac{1}{\sigma^2_{i_1,i_2}}+\frac{\sigma^{2(un)}_{i_1,i_2}(\mathbf{\Sigma}^{(old)})}{(\sigma^2_{i_1,i_2})^2} \\ &+ \alpha[(\ln\sigma^2_{i_1+1,i_2}+\ln\sigma^2_{i_1-1,i_2}+\ln\sigma^2_{i_1,i_2+1}+\ln\sigma^2_{i_1,i_2-1}-4\ln\sigma^2_{i_1,i_2}) \\ &\quad +\frac{1}{\sigma^2_{i_1,i_2}}(\sigma^2_{i_1+1,i_2}+\sigma^2_{i_1-1,i_2}+\sigma^2_{i_1,i_2+1}+\sigma^2_{i_1,i_2-1}-4\sigma^2_{i_1,i_2})]. \end{aligned} \tag{1.23}$$

Figure 4 shows the results of 1000 iterations of the EM algorithm using Goods's roughness penalty for $\alpha/(NM) = 0.002$ and $\alpha/(NM) = 0.005$.

*Silverman's Roughness*

Inspired by the work of Good and Gaskins [GG71], Silverman [Sil82] suggested alternative penalties that employ differential operators of the logarithm of the function. Like Good's roughness, Silverman proposed his penalty in the context of density estimation. We consider the special case

$$\int\int\left[\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial y}\right)\ln\sigma^2(x,y)\right]^2 dxdy. \tag{1.24}$$

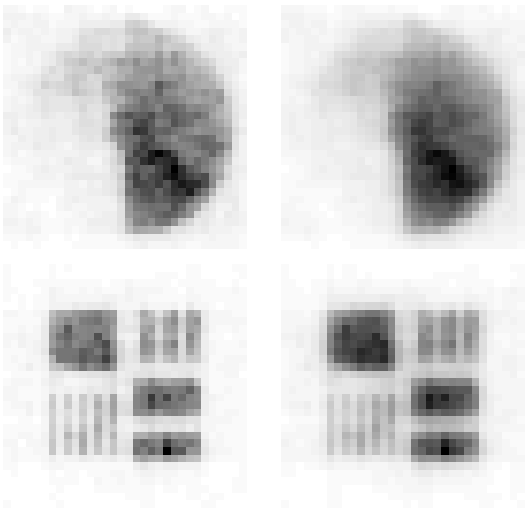

FIGURE 4. Results of 1000 iterations of the EM algorithm using Good's roughness penalty with $\alpha/(NM) = 0.002$ (left column) and $\alpha/(NM) = 0.005$ (right column).

In the context of optical astronomy, Molina and Ripley [MR89] suggest smoothing the logarithm of the image, noticing that "astronomers tend to look at the raw data on a logarithmic scale (by choosing contour levels in a geometric progression) except when looking at details at near background level."

Discretizing (1.24) yields

$$\Phi_S(f) = \sum_{i_1,i_2} [(\ln \sigma^2_{i_1+1,i_2} - \ln \sigma^2_{i_1,i_2})^2 + (\ln \sigma^2_{i_1,i_2+1} - \ln \sigma^2_{i_1,i_2})^2]. \quad (1.25)$$

Using this penalty requires solving the following set of nonlinear equations at each iteration of the EM algorithm:

$$\begin{aligned} 0 &= -\frac{1}{\sigma^2_{i_1,i_2}} + \frac{\sigma^{2(un)}_{i_1,i_2}(\mathbf{\Sigma}^{(old)})}{(\sigma^2_{i_1,i_2})^2} \quad (1.26) \\ &+ \frac{\alpha}{NM}\frac{2}{\sigma^2_{i_1,i_2}}[\ln \sigma^2_{i_1+1,i_2} + \ln \sigma^2_{i_1-1,i_2} + \ln \sigma^2_{i_1,i_2+1} + \ln \sigma^2_{i_1,i_2-1} - 4 \ln \sigma^2_{i_1,i_2}]. \end{aligned}$$

Figure 5 shows the results of 1000 iterations of the EM algorithm using Silverman's roughness penalty for $\alpha/(NM) = 0.002$ and $\alpha/(NM) = 0.005$.

*General Markov Random Fields*

Note that, when discretized, Good's and Silverman's roughness penalties can be thought of as inducing a prior with a nearest-neighbor Markov random field structure. A wide variety of such priors have been proposed for

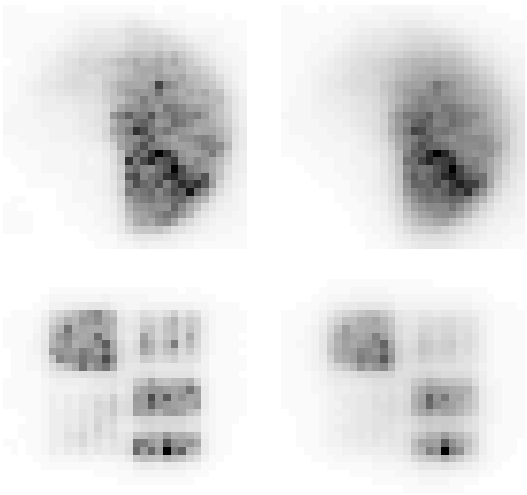

FIGURE 5. Results of 1000 iterations of the EM algorithm using Silverman's roughness penalty with $\alpha/(NM) = 0.002$ (left column) and $\alpha/(NM) = 0.005$ (right column).

image reconstruction [GR92] which could be tried here. For instance, the use of the square in (1.20) and (1.24) results in reconstructions that, while less noisy, have smoothed edges. Employing powers less than two results in a penalty that tends to smooth continuous regions while better preserving edges. For simple penalties employing the derivative of the function (instead of the square-root or the logarithm as above), this corresponds to a Markov random field with the popular generalized Gaussian distribution [BS93]. Good's roughness and Silverman's roughness could be generalized in a similar way. We leave this topic for future work.

### *4.3 Regularization via General Smoothing Steps*

Notice that in the method of penalties, a simple modification of the EM algorithm can be used to produce the penalized likelihood estimates; the resulting algorithm happens to amount to nonlinearly smoothing the result of the maximization step at each iteration before returning to the expectation step. In the fields of emission tomography and stereology, Silverman *et. al.* [SJWN90] have suggested experimenting with different kinds of smoothing steps. For arbitrary choices of smoothing, the resulting expectation-maximization-smoothing (EMS) algorithm will not, in general, correspond to a particular penalized likelihood method. In addition, it is difficult to determine exactly what such algorithms converge to, if they converge at all. However, considering the success illustrated in [SJWN90] for Poisson

intensity estimation, we are motivated to explore other kinds of smoothing in our radio astronomy application. One such example is shown in Figure 6, in which the smoothing step is a simple nearest-neighbor linear filter defined by

$$\begin{aligned}(Lf)_{i_1,i_2} &= f_{i_1,i_2} + \alpha[f_{i_1+1,i_2} + f_{i_1-1,i_2} + f_{i_1,i_2+1} + f_{i_1,i_2-1}] \\ &\quad + \frac{\alpha}{\sqrt{2}}[f_{i_1+1,i_2+1} + f_{i_1-1,i_2+1} + f_{i_1-1,i_2+1} + f_{i_1-1,i_2-1}],\end{aligned} \tag{1.27}$$

where $\alpha$ here is a parameter which controls the amount of smoothing.

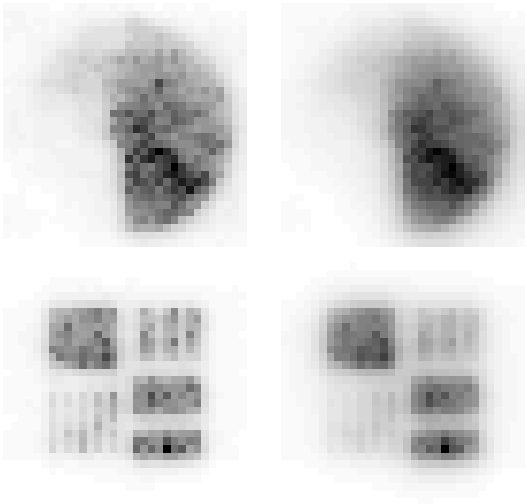

FIGURE 6. Results of 1000 iterations of an EMS algorithm using the linear smoothing step defined by (1.27) for $\alpha = 0.001$ (left column) and $\alpha = 0.003$ (right column).

Applying a linear smoothing step in the EM iteration for emission tomography has a natural relationship to minimizing a penalized likelihood in which the penalty is quadratic in the *square root* of the intensity (see [Nyc90] and Section 5 of [SJWN90]), yielding a rather unexpected connection to Good's roughness of Section 4.2.[2] Latham and Anderssen [LA94] provide some deep analysis of linear smoothing in the emission tomography iteration. We know of no equivalent existing analysis of adding a linear smoothing step to (1.10).

Eggermont and LaRiccia [EL95] propose smoothing by nonlinear operation $\mathfrak{N}$, constructed from a linear smoother according to $\mathfrak{N}f = \exp[L\ln(f)]$.

[2]Surprisingly, neither the authors nor the discussants of [SJWN90] refer to the work of Good and Gaskins or to its later application in emission tomography. The connection is explicitly noted by Eggermont and LaRiccia [EL95], however.

When applied to density or Poisson intensity estimation, the resulting EMS algorithm maximizes a particular functional they call a "modified loglikelihood". This contrasts with other EMS algorithms which do not necessarily correspond to minimizing some particular functional. To our knowledge, there is no parallel body of analysis of EMS algorithms for structured covariance estimation; it remains a wide-open area for future work.

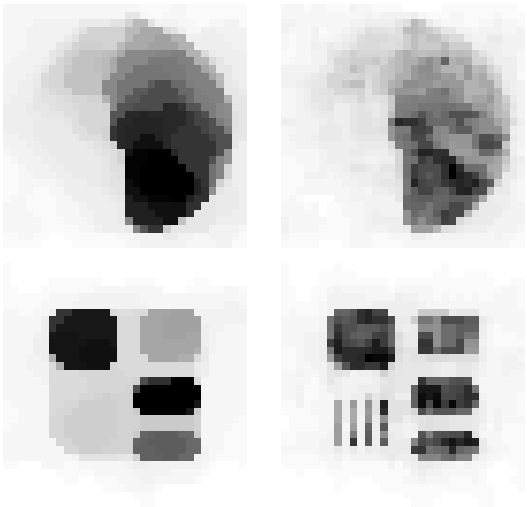

FIGURE 7. Left column shows results of 1000 iterations of an EMS algorithm using a nearest-neighbor median filter as the smoothing step. Right column shows results of median filtering 1000 iterations of the unconstrained EM algorithm.

For another example, the left column of Figure 7 shows the results of an EMS algorithm which uses a 3 by 3 median filter as the smoothing step. Observe how the EMS algorithm with median smoothing maintains sharp edges and yields a cartoon-like reconstruction; also note that this is *not* equivalent to median filtering the raw ML estimate of the unconstrained EM algorithm, as shown in the right column of Figure 7.

## 5 Directions for Future Work

The simulations presented in this chapter offer preliminary insight into the behavior of statistical estimation techniques for forming radio astronomy images. More detailed studies should be conducted, including studies with real data, which compare maximum-likelihood and maximum penalized-likelihood estimates with the results of traditional procedures such as the CLEAN algorithm and classical maximum-entropy algorithms.

Real radio astronomy measurements exhibit several phenomena which our statistical model currently does not account for. In some radio astronomy arrays, in order to reduce cost, the correlators are "hard wired" in such a way that only certain correlation measurements are made; thus certain elements of $\mathbf{S}$ may be missing. Before correlation, the raw data is often significantly truncated, sometimes as coarsely as one bit, which has an unpleasant effect on the correlation measurement (see Sections 8.3 through 8.5 of [TMS88]). It may be beneficial to explicitly model these effects and derive appropriate extensions to the EM algorithm presented here. There are also a host of calibration issues (particularly in Very Long Baseline Interferometry (VLBI), where the sensor array may span whole continents [FS89]) which we have neglected which could also be incorporated into the model and the algorithm. Leshem and van der Veen [LvdV00] discuss these calibration issues from a maximum-likelihood viewpoint.

One of the main difficulties with EM algorithms in general is their slow convergence rates. A variety of EM variants have been proposed which boast faster convergence. For instance, the Space-Alternating Generalized EM (SAGE) algorithm of Fessler and Hero [FH94] updates the parameters in groups instead of all at once; each group has its own associated *hidden data space*, which would be a complete data space if the remaining parameters were known. Schulz has formulated several SAGE algorithms [Sch97] for maximizing (1.4) which could be implemented for forming radio astronomy images. The one-dimensional experiments reported in [Sch97] demonstrate the greater likelihood change per iteration of the SAGE algorithm; however, that particular implementation requires around three times the amount of computation per iteration as the original EM implementation. The details of such timing analyses are highly platform dependent; one could imagine a specialized hardware architecture which could perform the SAGE iterations faster than the EM iterations, or vice-versa. In any particular application, one should compare various implementations to find the algorithm with best performance on some given hardware architecture for that particular instance. SAGE algorithms are more convenient for maximizing penalized likelihoods using Markov random field type penalties, because the SAGE recipe decouples the joint maximization step, which require solving sets of equations like (1.23) and (1.26), into a series of single-parameter maximizations.

An appealing aspect of the statistical formulation is that it allows the prediction of estimator performance via Cramer-Rao bounds. In addition to giving the scientist an understanding of the fundamental limits of the available instrumentation, it provides an appealing criteria for making design decisions, such as the placement of receiver locations, in various applications. (There has been a tremendous amount of work on studying such design choices [Cho72, Mat69] using other kinds of criteria.) For large images, inverting the Fisher information can become cumbersome; Hero and Fessler [HF94] propose an iterative algorithm for computing Cramer-Rao

bounds on parameter subsets which avoids explicit inversion of the Fisher information matrix via a complete/incomplete data formulation analogous to the EM algorithm.

*Acknowledgments:* I had the good fortune of being assigned Prof. Donald L. Snyder as my undergraduate advisor. I began working in his Electronic Systems and Signals Research Laboratory during the summer between my freshman and sophomore years. My first task was soldering the memory boards for a massively parallel computer. After that, Dr. Snyder kept giving me progressively more interesting tasks. At the start of my junior year, I began research in medical and astronomical imaging. This experience as an undergraduate hooked me on signal processing and ultimately inspired me to work towards my doctorate. Similarly, the example set by Dr. Snyder as a teacher and advisor encouraged me to pursue a career in academia.

I would like to thank Prof. Richard E. Blahut for detailed comments on several drafts of this manuscript.

This work was supported by a grant from DARPA under Contract F49620-98-1-0498, administered by AFOSR.

## 6 References

[Abl74] J.G. Ables. Maximum entropy spectral analysis. *Astronomy and Astrophysics Suppliment Series*, 15:383–393, 1974.

[AH91] U. Amato and W. Hughes. Maximum entropy regularization of Fredholm integral equations of the first kind. *Inverse Problems*, 7:793–808, 1991.

[BLW82] J.P. Burg, D.G. Luenberger, and D.L. Wenger. Estimation of structured covariance matrices. *Proc. of the IEEE*, 70(9):963–974, Sept. 1982.

[BM93] C.S. Butler and M.I. Miller. Maximum a posteriori estimation for SPECT using regularization techniques on massively-parallel computers. *IEEE Trans. on Medical Imaging*, 12(1):84–89, March 1993.

[BMMW94] C. S. Butler, M.I. Miller, T. R. Miller, and J. W. Wallis. Massively parallel computers for 3D single-photon-emission computed tomography. *Physics in Medicine and Biology*, 39:575–582, 1994.

[BS93] 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.

[Cho72] Y.L. Chow. On designing a supersynthesis antenna array. *IEEE Trans. on Antennas and Propagation*, 20:30–35, 1972.

[DJHS92] D.L. Donoho, I.M. Johnstone, J.C. Hoch, and A.S. Stern. Maximum entropy and the nearly black object. *Journal of the Royal Statistical Society B*, 54(1):41–81, 1992.

[EL95] P.P.B. Eggermont and V.N. LaRiccia. Maximum smoothed likelihood density estimation for inverse problems. *The Annals of Statistics*, 23:199–220, 1995.

[ES88] G.J. Erickson and C.R. Smith, editors. *Maximum-Entropy and Bayesian Methods in Science and Engineering*. Kluwer Academic, Dordrecht, 1988.

[FH94] J.A. Fessler and A.O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Trans. on Signal Processing*, 42(10):2664–2677, October 1994.

[Fom73] E.B. Fomalont. Earth-rotation aperture synthesis. *Proceedings of the IEEE*, 61(9):1211–1218, Sept. 1973.

[Fri72] B. R. Frieden. Restoring with maximum likelihood and maximum entropy. *J. Opt. Soc. Am.*, 62, No. 4:511–518, 1972.

[Fri91] B.R. Frieden. Some analytical and statistical properties of Fisher information. In *Stochastic and Neural Methods in Signal Processing, Image Processing, and Computer Vision*, volume SPIE Proc. 1569, 1991.

[FS89] M. Felli and R.E. Spencer, editors. *Very Long Baseline Interferometry: Techniques and Applications*. Kluwer Academic, 1989.

[FW78] B.R. Frieden and D. C. Wells. Restoring with maximum entropy. iii. Poisson sources and backgrounds. *J. Opt. Soc. Am.*, 68, No. 1:93–103, 1978.

[GD80] S.F. Gull and G.J. Daniell. The maximum entropy algorithm applied to image enhancement. *Proc. of the IEEE*, 5:170, 1980.

[GG71] I. J. Good and R. A. Gaskins. Nonparametric roughness penalties for probability densities. *Biometrika*, 58,2:255–277, 1971.

[GN99] A.K. Gupta and D.K. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC, 1999.

[Gol88] P.F. Goldsmith, editor. *Instrumentation and Techniques of Radio Astronomy*. IEEE Press, New York, 1988.

[GR92] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(3):367–383, March 1992.

[Gre81] U. Grenander. *Abstract Inference*. John Wiley and Sons, New York, 1981.

[H74] J.A. Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Suppliment Series*, 15:417–426, 1974.

[HF94] A. Hero and J.A. Fessler. A recursive algorithm for computing Cramer-Rao-type bounds on estimator covariance. *IEEE Trans. on Information Theory*, 40(4):1205–1210, July 1994.

[JD93] D.H. Johnson and D.E. Dudgeon. *Array Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1993.

[JM93] S. Joshi and M.I. Miller. Maximum a Posteriori estimation with Good's roughness for optical sectioning microscopy. *Optical Society of America A*, 10(5):1078–1085, May 1993.

[LA94] G.A. Latham and R.S. Anderssen. On the stabilization inherent in the EMS algorithm. *Inverse Problems*, 10:793–808, 1994.

[LvdV00] A. Leshem and A.-J. van der Veen. Radio astronomical imaging in the presence of strong radio interference. *IEEE Trans. on Information Theory*, 2000. Special issue on information theoretic imaging, to appear.

[Mat69] N.C. Mathur. A pseudodynamic programming technique for the design of a correlator supersynthesis array. *Radio Science*, 4:235–243, 1969.

[MB93] M.I. Miller and C.S. Butler. 3-D maximum a posteriori estimation for SPECT on massively-parallel computers. *IEEE Trans. on Medical Imaging*, 12(3):560–565, September 1993.

[Mea86] L.R. Mead. Approximate solution of Fredholm integral equations by the maximum entropy method. *Journal of Mathematical Physics*, 27:2903–2907, 1986.

[MM91] A.W. McCarthy and M.I. Miller. Maximum likelihood SPECT in clinical computation times using mesh-connected parallel computers. *IEEE Trans. on Medical Imaging*, 10(3):426–436, September 1991.

[MOS92] P. Moulin, J.A. O'Sullivan, and D.L. Snyder. A method of sieves for multiresolution spectrum estimation and radar imaging. *IEEE Trans. on Information Theory*, 38(2):801–813, March 1992.

[Mou93] P. Moulin. A wavelet regularization method for diffuse radar-target and speckle-noise reduction. *Journal of Mathematical Imaging and Vision*, 3(1):123–134, 1993.

[MR89] R. Molina and B.D. Ripley. Using spatial models as priors in astronomical image analysis. *Journal of Applied Statistics*, 16(2):193–206, 1989.

[MR91] M.I. Miller and B. Roysam. Bayesian image reconstruction for emission tomography incorporating Good's roughness prior on massively parallel processors. *Proc. of the National Academy of Sciences*, 88:3223–3227, April 1991.

[NN84] R. Narayan and R. Nityananda. Maximum entropy – flexibility versus fundamentalism. In J.A. Roberts, editor, *Indirect Imaging*, pages 281–290. Cambridge Univ. Press, 1984.

[NTE83] P.J. Napier, A.R. Thompson, and R.D. Ekers. The Very Large Array: Design and performance of a modern synthesis radio telescope. *Proceedings of the IEEE*, 71(11):1295–1320, Nov. 1983.

[Nyc90] D. Nychka. Some properties of adding a smoothing step to the EM algorithm. *Statistics and Probability Letters*, 9:187–193, 1990.

[O'S95] J.A. O'Sullivan. Roughness penalties on finite domains. *IEEE Trans. on Image Processing*, 4(9):1258–1268, September 1995.

[OSPM92] J.A. O'Sullivan, D.L. Snyder, D.G. Porter, and P. Moulin. An application of splines to maximum likelihood radar imaging. *International Journal of Imaging Systems and Technology*, 4:256–264, 1992.

[Pon80] J.E. Ponsonby. An entropy measure for partially polarized radiation and its application to estimating radio sky polarization distributions from incomplete "aperture synthesis" data by the maximum entropy method. *Monthly Notices of the Royal Astronomical Society*, 163:1973, 369-380.

[Poo94] V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, New York, NY, second edition, 1994.

[PS88] D.G. Politte and D.L. Snyder. The use of constraints to eliminate artifacts in maximum-likelihood image estimation for emission tomography. *IEEE Trans. on Nuclear Science*, 35(1):608–610, February 1988.

[PS91] D. G. Politte and D. L. Snyder. Corrections for accidental coincidences and attenuation in maximum-likelihood image reconstruction for positron-emission tomography. *IEEE Trans. on Medical Imaging*, 10(1):82–89, March 1991.

[RH60] M. Ryle and A. Hewish. The synthesis of large radio telescopes. *Monthly Notices of the Royal Astronomical Society*, 120:220–230, 1960.

[SB84] J. Skilling and R.K. Bryan. Maximum entropy image reconstruction: general algorithm. *Monthly Notices of the Royal Astronomical Society*, 211:111–124, 1984.

[Sch78] U.J. Schwarz. Mathematical-statistical description of the iterative beam removing technique (method CLEAN). *Astronomy and Astrophysics*, 65:345–356, 1978.

[Sch91] L.L. Scharf. *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Addison-Wesley, Reading, MA, 1991.

[Sch97] T.J. Schulz. Penalized maximum-likelihood estimation of covariance matrices with linear structure. *IEEE Trans. on Signal Processing*, 45(12):3027–3038, December 1997.

[Sil82] B. W. Silverman. On the estimation of a probability density function by the maximum penalized likelihood method. *The Annals of Statistics*, 10(3):795–810, 1982.

[SJWN90] B.W. Silverman, M.C. Jones, J.D. Wilson, and D.W. Nychka. A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography (with discussion). *Journal of the Royal Statistical Society B*, 52(2):271–324, 1990.

[SM85] D.L. Snyder and M.I. Miller. The use of sieves to stabilize images produced with the EM algorithm for emission tomography. *IEEE Trans. on Nuclear Science*, 32:3864–3872, 1985.

[SM91] D. L. Snyder and M. I. Miller. *Random Point Processes in Time and Space*. Springer-Verlag, 2nd edition, 1991.

[SMLTP87] D.L. Snyder, M.I. Miller, Jr. L.J. Thomas, and D.G. Politte. Noise and edge artifacts in maximum-likelihood reconstruction for emission tomography. *IEEE Trans. on Medical Imaging*, 6(3):228–237, 1987.

[SOM89] D.L. Snyder, J.A. O'Sullivan, and M.I. Miller. The use of maximum-likelihood estimation for forming images of diffuse radar-targets from delay-doppler data. *IEEE Trans. on Information Theory*, 35(3):536–548, May 1989.

[SSB79] J. Skilling, A. W. Strong, and K. Bennett. Maximum-entropy image processing in gamma-ray astronomy. *Monthly Notices of the Royal Astronomical Society*, 187:145–152, 1979.

[TMS88] A.R. Thompson, J.M. Moran, and G.W. Swenson, Jr.. *Inteferometry and Synthesis in Radio Astronomy*. John Wiley and Sons, New York, 1988.

[TT78] R. A. Tapia and J. R. Thompson. *Nonparametric probability density estimation*. Johns Hopkins University Press, Baltimore, 1978.

[WD77] S.J. Wernecke and L.R. D'Addario. Maximum entropy image reconstruction. *IEEE Trans. on Computers*, 26:351–364, 1977.