The Image Restoration/Reconstruction Problem

By R. C. PUETTER[1]

[1]Center for Astrophysics and Space Sciences,
University of California, San Diego
9500 Gilman Drive
La Jolla, CA, 92093-0111, USA
rpuetter@ucsd.edu

This lecture reviews the image restoration/reconstruction problem in its general setting. We first discuss linear methods for solving the problem of image deconvolution, i.e. the case in which the data is a convolution of a point-spread function and an underlying unblurred image. Next, non-linear methods are introduced in the context of Bayesian estimation, including Maximum-Likelihood and Maximum Entropy methods. Finally, the successes and failures of these methods are discussed along with some of the roots of these problems and the suggestion that these difficulties might be overcome by new (e.g. pixon-based) image reconstruction methods.

1. The Mathematics of the Image Reconstruction Problem

Image reconstruction is becoming a more and more common technique in optical and infrared astronomy. Of course, the problem of reconstructing images from Fourier domain data has be a long standing problem in radio interferometry (see Thompson *et al.* 1986; Cornwell and Perley 1991; Wohlleben *et al.* 1991). Indeed, in this discipline image reconstruction is a "bread-and-butter" application. The meaning of one's data simply cannot be interpreted from the raw data except in the very simplest of cases. In optical and infrared imaging, one distinguishes between image reconstruction and image restoration. As the names suggest, image restoration generally is a less drastic extrapolation from the data to the final image than is image reconstruction. Normally, image restoration refers to modifying data which already is in the form of an image and applying corrections to achieve an image which is superior in some way (e.g. has higher resolution or has reduced distortion). In the case of image reconstruction one is normally dealing with data that is not in the form of an image. The image is then literally reconstructed from this data. The example of radio interferometry is an excellent one here. In the case of radio interferometry the collected data is a sparsely sampled interferogram. While the data set (if complete) would fill a 2-dimensional image, the spatial relationships of features in the interferogram are very different from the spatial distributions of sources in the sky, and the image and the data are related to each other through a mathematical relationship that encodes the image in a complex manner.

While image restoration and reconstruction have many differences, both are examples of a larger class of problems called the "Inverse Problem". In the general inverse problem one seeks to evaluate, or estimate, the values of a certain set of parameters or functions from the known properties of other parameters or functions, e.g. one has the relationship:

$$\mathbf{L}\left(\{f_i\}, \{g_j\}\right) = 0 \quad , \tag{1.1}$$

where $\mathbf{L}$ is an operator (which can be quite general), the parameters (or functions), $\{f_i\}$, are sought, and the values of the parameters (or functions), $\{g_j\}$, are known.

In the context of the inverse problem, one often discusses the existence and uniqueness of the solution. In physical problems, the existence of a solution is assured, provided the problem is well posed. Furthermore, in the presence of noise, it is virually assured that the solution is not unique. After all, any two solutions differing only by properties within the noise limits (i.e. statistically indistinguishable solutions) are equally valid. A more interesting question is what are the properties of the solution space for various physical problems. The answer to this is rather complex, and varies from problem to problem. To gain some intuition, let us ask the question first in the image restoration framework. As already mentioned, any two statistically indistinguishable solutions are equally acceptable. That means that another solution can be generated from any known solution by adding a bit of noise. More precisely, to account for noise we would write:

$$\mathbf{L}\left(\{f_i\},\{g_j\}\right) = \mathbf{N} \quad , \tag{1.2}$$

where $\mathbf{N}$ is the noise which may be a scalar or multi-dimensional noise vector (note that this functional form may not be appropriate for all situations), then any solution, $\{f_i\}$, can generate new solutions, $\{f_i + \Delta f_i\}$, provided that

$$\mathbf{L}\left(\{f_i + \Delta f_i\},\{g_j\}\right) \approx \mathbf{N} \quad . \tag{1.3}$$

In fact, if $\mathbf{L}$ were a linear operator, we would talk about the null space of $\mathbf{L}$ and say that any operator contained in the null space could be added to any solution to construct new solutions. However, since in general we are not dealing with linear operators, this idea is not exact. Nonetheless, for the image restoration problem (which is often linear) it is clear that any small perturbation can be added to a known solution to produce another solution, provided that the perturbation produces an effect which is small relative to the noise.

We can also carry this exercise through for the radio interferometry problem. In this case, any solution can have added to it a function whose Fourier transform is small relative to the measured noise in the interferogram. Even more sinister for this case is the fact that huge regions of the Fourier plane are unsampled. Hence one is free to add functions with arbitrarily large Fourier components in these regions without violating the collected data, simply because no information is available at these frequencies. Clearly, in this situation there is a huge solution space. What is more, this solution space contains a wealth of unphysical solutions. Indeed, the typical solution is unphysical. It is clear in this case that in order to ensure the selection of a physically realistic solution, other conditions must be imposed, and indeed, this is the case—see below.

2. Linear Methods

While most inverse problems are non-linear in nature, an important linear inverse problem commonly occurs in astronomy. This is the case of image restoration of point-spread function (PSF) blurred object. In this case we have the situation:

$$D(\vec{x}) = \int dV_{\vec{y}}\; H\left(\|\vec{x}-\vec{y}\|\right) I(\vec{y}) + N(\vec{x}) \quad , \tag{2.4}$$

where D is the data collected, H is the PSF, I is the underlying, true, unblurred image, and N is the noise. The integral in equation (2.4) is over volume in image space and accounts for the general, multi-dimensional nature of this space. Further, we have assumed that the PSF is a function of the difference in position between $\vec{x}$ and $\vec{y}$, i.e. a convolution. Since equation (2.4) is a convolution, we can use the Fourier convolution theorem to solve the problem. In the absence of noise:

$$\begin{aligned} F(D) &= F(H * I) \quad , & (2.5)\\ &= F(H)F(I) \quad , & (2.6)\end{aligned}$$

$$I = F^{-1}\left(\frac{F(D)}{F(H)}\right) \quad , \tag{2.7}$$

where $F(f)$ is the Fourier transform of f. In the absence of noise, equation (2.7) provides an exact inversion of equation (2.4). This is a commonly used trick in many early attempts at image reconstruction. It is still often used for the reconstruction of bright objects in which the error due to noise is minimal. A good example of this is speckle image reconstruction (e.g. Ghez *et al.* 1993, Jones 1983) and Fourier transform spectroscopy (e.g. Serabyn and Weisstein 1995; Abrams *et al.* 1995; Prasad and Bernath; also see the review by Ridgway and Brault 1984). Unfortunately, because of the noise contribution to $D(\vec{x})$, we cannot exactly solve for the image in this manner. Fourier methods are notorious for their poor noise propagation properties. This can be understood through simple noise propagation arguments. As can be seen from equation (2.7), the image at each point will be obtained by calculating the Fourier transform of the data, dividing by the Fourier transform of the PSF and performing an inverse Fourier transform. Each of these operations requires a large number of adds and multiplies. If the original numbers are noisy, each add and multiply increases the noise. By the time the final answer is achieved, the propagated noise is quite large, giving rise to the characteristic "ringing" and spurious sources plaguing inverse Fourier methods. If this isn't bad enough, the PSF may have natural "zeroes" in its Fourier transform. These zeroes amplify the noise in the reconstruction. Furthermore, even if the PSF has no "natural" zeros in its Fourier transform, if the PSF is experiementally measured and suffers from noise, then many artificial zeroes may arise in the PSF's Fourier transform, causing noise problems.

Figure 1 presents a comparison of the performance of direct Fourier deconvolution to two non-linear techniques, i.e. straight GOF deconvolution and pixon-based deconvolution—see Lecture 3 for a description of pixon-based methods. For the purposes of this comparison, a mock data set was constructed. The true underlying image is displayed at the top of Figure 1. This is the goal image of the reconstructions. The input data was created from the true image by convolving it with a Gaussian PSF with FWHM=8 pixels. To this image Gaussian noise was added to give a peak signal-to-noise ratio of 30. This was the input data for each of the deconvolutions. As can be seen from the reconstructions, the direct Fourier inverse is by far the poorest. There is a great deal of noise in the reconstruction and the achieved resolution is poor. In addition, the residual errors are quite large. By comparison the non-linear methods give better results with excellent residuals and considerably better resolution. Note, however, that in the straight GOF reconstruction a number of spurious sources have been created. This is indicative of overfitting of the data. By far the best reconstruction is given by the pixon-based technique. Here the residuals are excellent, the resolution is high, and there are no spurious sources.

In summary, Fourier methods provide a simple inversion technique for convolution problems. However, these techniques are only effective if the data are virtually noise free and the PSF is well known and noise free. Because of these difficulties, most modern efforts in image reconstruction methods have turned to non-linear methods. Such an approach holds significant benefit for controlling the propagation of noise in the solution.

3. Non-Linear Methods

Most non-linear methods can be interpreted in terms of a Bayesian estimation scheme in which the hypothesis sought is in some sense the most probable. To derive a suitable goal function with which to judge the relative merits of various hypotheses, Bayesians use conditional probabilities to factor the joint probability distribution of D, I, and M, i.e. $p(D, I, M)$, where D, I, and M are the data, unblurred image, and model respectively—

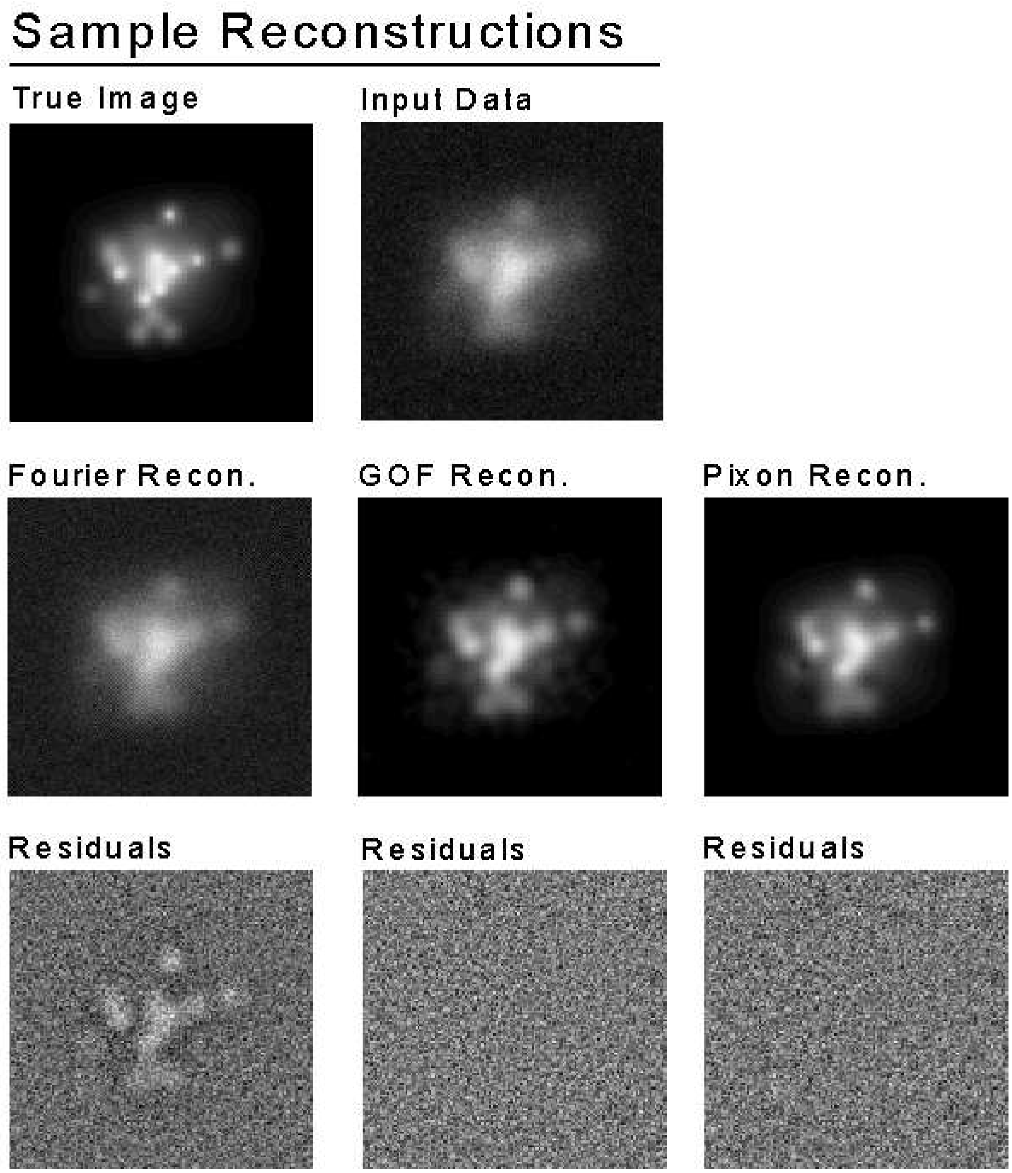

FIGURE 1. Comparison of image reconstruction techniques for a mock data set. Shown at the top is the true, underlying, high resolution, noise-free image. This has been convolved with a Gaussian PSF with FWHM=8 pixels and Gaussian noise added to make the input data image. The peak signal-to-noise ratio in the data is about 30. Shown at the bottom of the figure are reconstruction results for a direct Fourier deconvolution, a pure GOF deconvolution, and a pixon-based deconvolution. Beneath each deconvolution are the resulting residual errors, i.e. the result of subtracting the reconstruction convolved with the PSF from the input data.

we now explicity assume that the data plays the role of $\{g_j\}$ in equation (1.3) and the image plays the role of $\{f_i\}$. The model, M, includes all aspects of the relationship between D and I such as the physics of the image encoding process, the details of the measuring instrument, e.g. pixel size, noise properties, etc., and the mathematical method of modeling the data, e.g. that equation (1.3) might be discretized, etc. As we shall see, all aspects of the model are important and can affect the quality of the solution to the inverse problem.

To derive equations useful for inverting equation (1-3), Bayesians factor $p(D, I, M)$ as:

$$
\begin{aligned}
p(D, I, M) &= p(D|I, M)p(I, M) \quad , && (3.8)\\
&= p(D|I, M)p(I|M)p(M) \quad , && (3.9)\\
&= p(I|D, M)p(D, M) \quad , && (3.10)\\
&= p(I|D, M)p(D|M)p(M) \quad , && (3.11)\\
&= p(M|D, I)p(D, I) \quad , && (3.12)\\
&= p(M|D, I)p(D|I)p(I) \quad , && (3.13)
\end{aligned}
$$

where $p(X|Y)$ is the probability of X given that the value of Y is known. By equating various terms, these equations give rise to the formulae

$$p(I|D, M) = \frac{p(D|I, M)p(I, M)}{p(D|M)} \propto p(D|I, M)p(I|M) \quad , \tag{3.14}$$

$$p(I, M|D) = \frac{p(D|I, M)p(I|M)p(M)}{p(D)} \propto p(D|I, M)p(I|M)p(M) \quad . \tag{3.15}$$

The formula of equation (3.14) is the typical starting place for Bayesian methods. This equation essentially assumes that the appropriate model, M, is known and will not be varied during the reconstruction. This assumption is made explicit in the proportionality of equation in (3.14) where the term $p(D|M)$ is dropped (it is assumed to be constant—D and M are not varied during the reconstruction). Our preferred formulation, however, is given by the formula of equation (3.15). Here we do not assume that all of the parameters associated with the model belong to the so called "nuisance parameters", i.e. parameters which might require estimation, but which are of no interest to the scientist. In pixon-based reconstruction, however, the model parameters associated with the local smoothness scale of the image (see later lectures) are extremely important.

The significance of the terms on the far right hand side of equations (3.14) and (3.15) is readily understood. The term $p(D|I, M)$ is a goodness-of-fit (GOF) quantity, measuring the likelihood of the data given a particular image and model. The terms, $p(I|M)$ and $p(M)$, are "priors", and incorporate our prior knowledge (or expectations) about the measurement situation or the nature of suitable selections for I and M. The term "prior" is used since each of these terms make no reference to the data, D, and hence can be decided on *a priori*, i.e. before the act of making the measurement.

In GOF (or maximum likelihood) image reconstruction, $p(I|M)$ and $p(M)$ are assumed to be constant, i.e. there is no prior bias concerning the image or parts of the model that might be varied. The standard choice for $p(D|I, M)$ is to use $p(D|I, M) = (2\pi\sigma^2)^{1/2n_{pixels}} \exp\left(-\chi^2/2\right)$, i.e. the joint probability distribution for n_{pixels} independent pixels with normally distributed (Gaussian) noise. This assumes that (1) within the chosen space of image/model pairs, a statistically acceptable fit to the data is obtainable, and (2) that near this solution, image/model space uniformly fills data space so that the resulting probability distribution of data realizations is dominated by the probability density of the Gaussian distributed measurement noise, i.e.

$$
\begin{aligned}
p(\text{Blurred Image+Noise}|I, M) &= p(\text{Blurred Image}|\text{Noise}, I, M)p(\text{Noise}|I, M) && (3.16)\\
&\cong \mathit{const} \times p(\text{Noise}|I, M) \quad . && (3.17)
\end{aligned}
$$

The "constant" $p(\text{Blurred Image}|\text{Noise}, I, M)$ is then generally dropped from the expression. This is generally not too bad an assumption, but it is by no means guaranteed, especially if the chosen space of image/models is selected to be restrictive in some way.

In Maximum Entropy (ME) image reconstruction, the image prior is based upon "phase space volume" or counting arguments and the prior is expressed as $p(I|M) = \exp(\alpha S)$, where S is the entropy of the image and α is an adjustable constant which weights the relative importance of the GOF and image prior. Again, usually the model is considered to be known and held fixed, i.e. $p(M) = constant$, and the usual assumption is to choose$p(D|I, M) = (2\pi\sigma^2)^{1/2n_{pixels}} \exp\left(-\chi^2/2\right)$. While many specific formulations for α and S appear in the literature (Kikuchi and Soffer 1976; Bryan and Skilling 1980; Narayan and Nityananda 1986; Skilling 1989), all ME methods capitalize on the virtues of incorporating prior knowledge of the likelihood of the image into the image restoration algorithm.

Each of the equations (3.14)-(3.15) allow the calculation of the Maximum *A Posteriori* (M.A.P.) image (or image/model pair), i.e. the image (image/model) which maximizes $p(I|D, M)$ [or $p(I, M|D)$]:

$$I_{MAP} = \arg\left\{\max_I \left[p(I|D, M)\right]\right\} \quad , \tag{3.18}$$

$$\{I_{MAP}, M_{MAP}\} = \arg\left\{\max_{I,M} \left[p(I, M|D)\right]\right\} \quad . \tag{3.19}$$

The M.A.P. image or image/model, of course, is only one possible choice of what one might select as the "best" solution. This choice selects the arguments that maximize the posterior probability density function. Another sensibile choice, however, would be to choose the mean image or image/model pair:

$$< I > = \int dM\; dD\; I\; p(I|D, M) \quad , \tag{3.20}$$

$$< I, M > = \left(\int dD\; I\; p(I, M|D), \quad \int dD\; M\; p(I, M|D) \right) \quad . \tag{3.21}$$

4. The Bayesian Embodiment of Occam's Razor

While most scientists have heard of Occam's Razor, and most of the rest have unknowingly embraced the concept as a working rule of thumb in their research, it is not widely understood that Occam's razor can be justified on solid probabilistic grounds. (The principle of Occam's Razor states that the best hypotheses, or explanation, of a data set is the simplest one.) One of the best discussions of the Bayesian embodiment of Occam's Razor is that presented by David MacKay (MacKay 1994). Our discussion shall parallel MacKay's discussion.

In order to judge the relative merits of two hypotheses, H_1 and H_2, for explaning the data set D, Bayesian's would form the relative probability ratio:

$$\frac{p(H_1|D)}{p(H_2|D)} = \frac{p(H_1)}{p(H_2)} \frac{p(D|H_1)}{p(D|H_2)} \quad . \tag{4.22}$$

The first ratio involves hypothesis priors, i.e. hypothesis probability densities with no functional dependence on the data. These can be used to express our prior conceptions of the relative likelihood of hypothesis H_1 relative to H_2. As pointed out by MacKay, this corresponds to the motivations behind Occam's razor. However, it is unnessary to make any prior assumptions about the relative probabilities $p(H_1)$ and $p(H_2)$. Equation (4.22) will still give rise to an Occam's Razor like effect. This is because the simpler hypotheses, say H_1, make the most precise predictions. Complex hypotheses embrace a much larger data space and hence spread their predictive probability $p(D|H_2)$ more thinly over this larger data space. Thus, in the case in which both hypotheses H_1 and H_2 are equally compatible with the data, the simpler hypothesis H_1 will be favored by equation (4.22), without having to express any subjective dislike for one hypothesis relative to the other. Of course even if one assigns an equal prior probability to certain aspects of the hypothesis, e.g. the model in image reconstruction, one can usually make sensible arguments for appropriate priors for the other aspects of the hypothesis, e.g. the image prior. For example, in the image reconstruction case, the hypothesis for a given data set is the pair (I, M). The prior is thus $p(I, M) = p(I|M)p(M)$. So even if one expresses no prior bias for a given model, e.g. $p(M) = constant$, one might easily assign a sensible prior probability distribution for $p(I|M)$, e.g. a ME prior. Such an assignment also acts in the cause of simple models, since any sample state drawn from a simple model typically has a higher probability simply because it represents a larger fraction of all possible states. Thus at every turn probability theory favors Occam's Razor.

5. Successes and Failures of Classical Methods

The very fact that image restoration/reconstruction has been performed continually for many years is testimony to its basic success as a method. For PSF blurring problems, its primary advantage is its ability to provide enhanced resolution. As discussed above, for linear problems, Fourier methods are able to provide such enhancement if noise is absent or inconsequential. In the low noise case, only digitization noise [i.e. the fact that the data or equation (1.1) is descretized in some manner for computer solution] and natural "zeroes" in the Fourier transform of the PSF limits our ability to achieve infinite resolution. While Fourier techniques are widely used for high SNR data, this techniques fails badly for moderate to low SNR data.

Non-linear methods provide an improvement over linear methods in moderate to low SNR situations. Unlike Fourier methods, non-linear methods acheive significantly greater control of noise propagation. This is done by insisting (through the GOF criterion) that the modeled data (e.g. the data predicted by the inferred image and model) match the measured data within the uncertainty specified by the noise. This is quite different from Fourier methods were no suchdemand is made. Of course, insisting that the predicted data "fits" the measured data within the noise limits does not guarantee that the inferred image is without significantly larger noise. Indeed, as discussed above, each image reconstruction problem can have a significant "null space'. [Quotes are used around "null space" since there need not be a strict null space for the data encoding problem under consideration, but large Δf_i relative to f_i may exist that still satisfy equation (1.3).]

Common problems with GOF (or Maximum Likelihood) and ME methods are overfitting of the data, the production of signal correlated residuals, and spurious sources. Overfitting of the data occurs when the GOF criterion is driven below expected values. For example, suppose an image is acquired with a $n \times n$ pixel CCD system. Further suppose that the CCD plate scale finely samples the PSF (say, m pixels across the PSF FWHM). Normally GOF reconstruction tries to drive the value of χ^2 to zero to obtain the best fit. Indeed, this is the proper prescription for obtaining the best estimate of n^2 independent pixel values and provides a consistent and unbiased estimate of the residuals. Consequently this proceedure would provide an excellent image estimate if a large sample of images were taken, the reconstructions performed, and the average value of the reconstructed image evaluated. However, this is usually not the situation one is faced with. Normally one wishes to estimate the best image from a single data frame. In this case, attempting to drive the value of χ^2 to zero often causes overfitting of the data. After all, the expected value of χ^2 is not zero for n^2 independent, Gaussian distributed variables, but is equal to $n^2 - 2$ (= peak, or mode, of the χ^2 distribution). Another way of stating this is that if one does a perfect job of estimating the underlying image, after convolving with the PSF and subtracting the result from the data, one expects to see a random noise frame. The expected (modal) value of χ^2 for this noise frame is not zero, but is $n^2 - 2$ with a standard deviation of $\sigma_{\chi^2} = \sqrt{2n^2}$. In this case, the likelihood function for the data (given the image and mode) is provided by the χ^2-distribution for n^2 independent variables.

As already mentioned, the production of signal correlated residuals and spurious sources are common problems with GOF and ME methods. Both problems are related to using an inappropriate number of degrees-of-freedom (DOFs) to fit the data and consequently are largely overcome by pixon-based methods—see later lectures in this series. Briefly, however, signal correlated residuals are produced by a "diluted" GOF criterion. Suppose we have imaged a galaxy in a large CCD frame. Assume that the CCD has dimensions 1024×1024 pixels. Further suppose that the galaxy only fills the central 100×100 pixels. If the astronomer were to use the entire data frame in his reconstruction and chose to represent the image as a rectangular grid of numbers with the same spatial frequency as the data (as is commonly done), then the reconstruction would use roughly 10^6 DOFs. If a standard GOF criterion were used, then one would stop the iterative procedure when there were less than 10^3 residuals larger than 3σ (0.1% of the residuals are 3σ or larger in a population of Gaussian distributed variables). Since iterative procedures that minimize GOF functions normally spend most of their time fitting the bright sources, they will be working on adjusting the bright source levels when the stopping criterion is met, resulting in the vast majority of the large residuals lying under the bright sources.

The origin of spurious sources in standard image reconstruction techniques is even easier to understand. Using the example above, it should be clear to most readers that the 10^6 DOFs used in the reconstruction are probably far more than are needed to describe the information present in the data. In fact, the data probably is inadequate to constrain the values of this many DOFs. Hence the vast majority of these DOFs are free to produce whatever bumps and wiggles they like, with the only requirement being that after they are smoothed by the PSF [or acted on by L in equations (1.2)] they average to zero within the noise. Hence their amplitude can be very large as long as they have a spatial scale small relative to the PSF and are arranged in such a manner that hills and valley tend to cancel each other out. Since this large number of unconstrained DOFs represents a huge phase space and the typical member has numerous bumps and wiggles, spurious sources are guaranteed.

One of the main motivations for developing Maximum Entropy methods was to help overcome the spurious source problem. After all, the ME prior acts to make unsmooth images undesireable. However, ME methods in their zeal to flatten spurious sources, ME methods can over flatten the true sources. This affects the ability of ME mthods to perform statistically unbiased photometry. In fact, while ME methods are significantly more immune to spurious sources (they more effectively reject situations with large adjancent hills and valleys than GOF methods) they still tend to systematically underestimate source brightnesses (Sibisi 1990; Cohen 1991) because of using too many DOFs in the reconstruction.

6. The Promise of Pixon-Based Methods

Most of the problems facing standard GOF and ME methods are over come by pixon-based methods. Pixon-based methods are a generalization of ME methods and perform the Bayesian estimation problem in a more natural language. The pixon language is information-based. Consequently it concisely and accurately describes the key variables in the problem and avoids the use of too many DOFs in the reconstruction. Because the number of DOFs is tuned to the number required to accurately describe the image, problems with stopping the reconstruction because a "diluted" GOF criterion is used are avoided. In addition, spurious source production are avoided since the pixon-basis automatically rejects spurious sources as being unneeded in the reconstruction. The generalization of the pixon to arbitrary data is the *informaton*, i.e. a quanta of information—see later lectures in this series, and has a very large range of application to a variety of problems. The informaton is directly related to a data set's Algorithmic Information Content—see Lecture 2—and is the appropriate coordinate system for the general inverse problem and optimal data compression. The value of using a particular language for image reconstruction will be motivated in Lecture 2 of this series, and pixon-based methods will be developed in Lecture 3.

REFERENCES

ABRAMS, M. C., DAVIS, S. P., RAO, M. L. P., ENGLEMAN, R., AND OTHERS 1995, High ResolutionFourier Transform Spectroscopy of the Meinel System of OH *Ap. J. (Suppl.)***93**, 351-395.

BRYAN, R. K., AND SKILLING, J. 1980, *M.N.R.A.S.***191**, 69.

COHEN, J. G. 1991, In *Maximum Entropy and Bayesian Methods*, ed.s W. T. Grady, Jr. and L. H. Schick. Kluwer Academic Publishers

CORNWELL, T. J., AND PERLEY, R. A., EDS. 1991, *Radio Interferometry: Theory, Techniques, and Applications*, (San Francisco: Astronomical Society of the Pacific)

GHEZ, A. M., NEUGEBAUER, G., AND MATHEWS, K. 1993, The Multiplicity of T Tauri Stars in the Star Forming Regions Taurus-Auriga and Ophiychus-Scorpius: A 2.2 Micron Speckle Imaging Survey *A. J.***106**, 2005-2023.

JONES, R. 1983, *Holographic and speckle Interferometry: A Discussion of the Theory, Practice and Application of the Techniques*. Cambridge University Press.

KIKUCHI, R., AND SOFFER, B. H. 1976, In *Image Analysis and Evaluation,* Society of Photographic Scientists and Engineers, Toronto, Canada, July 1976, 95.

MACKAY, D. J. C. 1994 in *Models of Neural Networks III*, eds. E. Domany, J. L. van Hemmen, and K. Schutten, Chapter 6, Springer-Verlag.

NARAYAN, R. AND NITYANANDA, R. 1986, Maximum Entropy Image Restoration in Astronomy, *Ann. Rev. Astron. & Astrophys.***24**, 127.

PRASAD, C. V. V., AND BERNATH, P. F. 1994, Fourier Transform Spactroscopy of the Swan (D(3)PI(G)-A(3)PI(U)) System of the Jet-Cooled C2 Molecule, *A. J.***426**, 812-821.

RIDGWAY, S. T., AND BRAULT, J. W. 1984, Astronomical Fourier Transform Spectroscopy Revisited, *Ann. Rev. Astron. & Astrophys.***22**, 291-317.

SERABYN, E., AND WEISSTEIN, E. W. 1995, Fourier Transform Spectroscopy of the Orion Molecular Cloud Core, *Ap. J.***451**, 238-251.

SIBISIS, S. 1990 in *Maximum Entropy and Bayesian Methods*, ed. J. Skilling. Kluwer Academic Publishers.

SKILLING, J. 1989, Classic Maximum Entropy, In *Maximum Entropy and Bayesian Methods*, ed. J. Skilling, p. 45. Kluwer Academic Publishers.

THOMPSON, A. R., MORAN, J. M., AND SWENSON, G. W. JR. 1986, *Interferometry and Synthesis in Radio Astronomy*. Wiley.

WOHLLEBEN, R., MATTES, H., AND KRICHBAUM, TH. 1991, *Interferometry in Radio Astronomy and Radar tecniques*, (Dordrecht; Boston: Kluwer Academic Publishers).