Methods of Deconvolution

Jörg Pfleiderer
Institut für Astronomie der Universität
Innsbruck, Austria

Abstract

We describe the mathematical problem of deconvolution and discuss 3 classes of methods to solve it: (1) methods working in Fourier space. (2) methods working in image space, with smoothness constraints, and (3) methods for image improvement without true deconvolution. The highest resolution can be achieved by class-2 methods.

1 Introduction

Deconvolution as a means of improving the resolution of instruments is well established for more than a century, Lord Rayleigh being the first to treat the problem mathematically. Soon it was recognized that the most reliable method of resolution increase was to use better instruments. The interest in mathematical methods has been revived in recent years (see, *e.g.*, Pfleiderer and Reiter 1982). This is because the performance of observing instruments has reached a maximum in many fields, and cannot be improved much by building larger or better instruments (or only with excessively high costs). So one should extract as much information from the data as possible. Also, the increase in computer performance permits the use of methods with large numerical effort. One of the most favoured methods presently used is the maximum entropy method MEM. For radiointerferometric observations, the method most often applied is CLEAN, of which several improvements over the original version of the seventies (Högbom 1974) exist.

In Sect. 2, we state the mathematical problem. Sects. 3 - 5 describe classes of deconvolution methods which are discussed in Sect. 6.

2 The deconvolution problem

The convolution/deconvolution equation in image space is, in discrete form,

$$f_i = \sum_j g_j h_{ij} + e_i \ , \tag{1}$$

where i and j are pixel numbers (which, in a two-dimensional problem, would each have two components), f is the observable signal (data), h the point spread function PSF or beam (depending, in a true convolution, only on the distance between pixels i and j), e the error, and g the true signal, approximated by a set of point sources at pixels j. The data pixels $i = 1, \ldots, I$ and the signal pixels $j = 1, \ldots, J$ need not necessarily have the same size. It is by no means necessary at this stage to assume a special form of the error distribution, e.g., that the noise data e_i are independent from each other or from the source distribution or the data. In the convolution, the signal g is convolved with the PSF h and noise e is superimposed to give the data f. In the deconvolution, the solution of the problem consists of finding a model set of signals $\{m_j\}$ such that the residuals

$$r_i := f_i - \sum_j m_j h_{ij} \tag{2}$$

are left equal to the errors. The true solution $r_i = e_i$ cannot be found because e is not known. A possible or acceptable solution is one for which the statistics of the set $\{r_i\}$ is sufficiently similar to what is expected for the error distribution $\{e_i\}$.

The Fourier transform of Eqn. 1 is

$$F_k = G_k \cdot H_k + E_k \ , \tag{3}$$

where the index k stands for a spatial frequency u_k or frequency range $[u_k, u_k + du_k]$. F, G, H, and E are the coefficients of the transform of f, g, h, and e, respectively.

Consider a point source of unit strength at pixel j, with Fourier transform $G_k(j)$. Then

$$H_{kj} := G_k(j) \cdot H_k \tag{4}$$

is the Fourier response to that source. From the linearity theorem, it follows that the response to a source of strength g_j is $g_j H_{kj}$, and the response to a set of point sources is $\sum_j g_j H_{kj}$. That is, another and entirely equivalent form of Eqn. 3 is

$$F_k = \sum_j g_j H_{kj} + E_k \ , \tag{5}$$

which comprises, if seen as a deconvolution problem, a linear system of K equations

$$R_k = F_k - \sum_j m_j H_{kj} \ , \quad k = k_1, \ldots, k_K \tag{6}$$

for the J unknowns m. Here, the R's are the Fourier residuals. Again, the true solution $R_k = E_k$ cannot be found but only possible ones in which the statistical properties of the residuals are accceptable. Accordingly, there are infinitely many possible solutions, both for Eqn. 2 and Eqn. 6.

3 Fourier methods

They work completely in Fourier space and use Eqn. 3. Many PSFs, in particular Gaussian ones, have Fourier coefficients which are noticeably different from zero only within a finite frequency interval $[-u_c, u_c]$, thus comprising a low pass filter. A good example is a single-dish radio observation. Higher frequencies are not transmitted by the instrument. Low frequencies are well transmitted while intermediate frequencies are partially filtered out.

Bracewell and Roberts (1954) introduced, for that case, the so-called principal solution

$$G_k = F_k/H_k \quad \text{for} \quad |u_k| < u_c \; ; \tag{7a}$$

$$G_k = 0 \qquad \text{otherwise} . \tag{7b}$$

It tends to show oscillations of frequency $\sim u_c$ in the image-space solution (Fourier transform of G) remindful of interference patterns, and negative values around pointed positive ones - a kind of Gibbs phenomenon. This is why a constraint of non-negativity can much improve the solution.

For a more complicated PSF, the principal solution can be generalized to be

$$G_k = F_k/H_k \quad \text{for} \quad |H_k| > C \; , \tag{8a}$$

$$G_k = 0 \qquad \text{otherwise} , \tag{8b}$$

where C is a suitably small constant. Eqns. 7a and 8a reverse the partial filtering of coefficients caused by low values of H and are, therefore, called "inverse filtering".

The sharp cut-off can be somewhat smoothed. In terms of a least-squares optimization, the best approach is the Wiener filter (Wiener 1942, Helstrom 1967) which was introduced into astronomy by Brault and White (1971):

$$G_k = \frac{F_k H_k^\star}{H_k H_k^\star + \Phi_k} , \tag{9}$$

where Φ is the ratio of the spectral densities of the signal and the noise. Φ is not known but can be reasonably estimated.

Most Fourier methods have the disadvantage that they put $G = 0$ for those frequencies which were lost in the process of observation, zero being the most non-committal default value. They are, therefore, unable to recover very steep features and to avoid Gibbs phenomena.

Some iterative methods work in image space but nevertheless are Fourier methods inasfar as they converge to the inverse-filter solution (Frieden 1975). In order to avoid explosion of high-frequency components, the iteration must be stopped before convergence at some optimum point, or some other measure must be introduced (Jansson *et al.* 1970). We mention two methods:

Van Cittert's (1931) algorithm can (with use of the same pixels i for data and signal) be written as

$$m_i(0) = 0 \; , \quad m_i(n+1) = m_i(n) + r_i(n) \; , \tag{10}$$

where n is the iteration number. The method has been used, *e.g.*, in solar physics (Wittmann 1971), in molecular beam scattering (Siska 1973), in geophysics (Ioup and Ioup 1983).

The algorithm of Lucy (1974) is, actually, not a pure Fourier method. It iteratively estimates the inverse beam $k := h^{-1}$ which reproduces the signal when the data is convolved with it:

$$m_j(n+1) = \sum_i f_i k_{ji}(n) \; , \quad k_{ji}(n) = \frac{m_j(n) h_{ij}}{\sum_l m_l(n) h_{il}} \; . \tag{11}$$

With the improvement of other methods and computer availability, it seems that pure Fourier methods are somewhat outdated and will be used only in some special applications (Subrahmanya 1980).

4 Image-space methods with smoothing constraint

As long as $J \geq I$ or $J \geq K$, Eqns. 2 or 6 can be solved for exactly vanishing residuals. This noise-fitting procedure would give an excessively oscillating solution, coinciding with the inverse-filter solution without cut-off: The high frequencies of the solution, after being strongly damped by the convolution with the beam, still have to reproduce the finite high-frequency amplitudes of the noise. Obviously, such solutions are not acceptable.

In order to find an acceptable solution, one has to use Eqn. 2 or 6 in a slightly modified form: The lefthand sides are first artificially neglected in order to have a well-defined linear system of equations but then the system is not solved exactly but only approximately. For selecting one of the infinite number of acceptable solutions, one has to introduce a constraint. Since small-scale structure in the true signal is damped out by the convolution, it cannot be recovered from the data with any certainty. Therefore, the constraint should suppress such features, selecting essentially the smoothest solution compatible with the data. A non-negativity constraint is also quite helpful (Biraud 1969) but is, depending on the problem, not always possible.

There are no standard methods for solving a linear system of equations approximately. Also, a nonlinear constraint destroys the linearity. Therefore, each method uses a different kind of iterative algorithm, adapted to the constraint. In most methods, the underlying philosophy is a least-squares fit. Then the residuals should resemble a normal distribution with average zero and a given variance which is known or can be estimated from the measuring error. For example, MEM deconvolves to a certain value of χ^2 or to a more refined error statistics (E^2 distribution: Bryan and Skilling 1981; position-independent distribution: Reiter and Pfleiderer 1986).

While the algorithm actually used is generally only of marginal influence on the result, the choice of a good smoothness constraint is essential. Some authors stress the necessity for a convincing constraint philosophy, as non-committance or simplicity, while others consider every constraint which provides a sufficiently smooth result as a good one (Nityananda and Narayan 1982).

Oscillatory solutions will always result in large values of the second derivative of the source distribution. Minimizing the second derivative in a least-squares sense is, therefore, a good constraint and has been used by a number of authors (Phillips 1962, Tikhonov 1963, Twomey 1963, Turchin and Turovtseva 1974, Tikhonov and Arsenin 1977, Subrahmanya 1980, Basistov *et al.* 1979, Jonas 1985). That this constraint seems to be not much used nowadays is probably not a result of the constraint being inferior but rather a result of the numerical algorithms used being inferior.

Another feature of oscillatory solutions is that some pixels have unnecessarily large content, which is recognizable, for example, by a large square. Minimization of the sum of squared pixel contents will again avoid such cases and thus produce a smooth solution. This is the constraint of the so-called Smoothness-Stabilized CLEAN or SSC (Cornwell 1983). The method can also be described as an (unconstrained) deconvolution with a modified beam which is the PSF with a central peak added. This is why it is also called Prussian-Helmet CLEAN or Prussian-Hat CLEAN. Other powers of the pixel contents than the second have also been discussed. For example, a maximization of the sum of square- rooted pixel contents is about as good.

Similarly, one can consider differences in the contents of adjacent or nearby pixels as unsmooth and try to minimize those differences as a function of the distance of pixels. This constraint can be formulated as giving a minimum of information on small-scale structure which was partly or wholly lost in the data-collecting process by the smoothing effect of the convolution with the PSF, hence the name Minimum Information Method MIM (Pfleiderer 1985, 1988). The derivation of a corresponding expression for structural information from general premises, such as invariances, will be given in a forthcoming paper (Pfleiderer, in preparation). The method is related to SSC. In particular, it also uses a kind of Prussian Helmet PSF.

Maximum entropy is characterized by a different approach to the question what "structure" is and what kind of structure should be suppressed. The philosophy of MEM has been described in a large number of papers (see, *e.g.*, Jaynes 1957, Frieden 1972, Ables 1974). The original main disadvantage of MEM, *viz.* the large size of the computer program which made MEM inaccessible to the average user, is now much eased by the availability of more compact programs and larger computers.

MEM has been the most successful deconvolution routine so far, with applications to a wide field of problems, as main beam deconvolution, photography (Bryan and Skilling 1981), interferometry (Wernecke 1977, Gull and Daniell 1978, Nityanda and Narayan 1982, Sanromà and Estalella 1984), incomplete data (Gull and Daniell 1978), spectral analysis and time series (Jensen and Ulrych 1973, Komesaroff *et al.* 1981), computer tomography, seismology.

5 CLEAN

This method, dating back to Högbom (1974), was specially devised for handling incomplete interferometric radio data. In image space, the incompleteness of Fourier data can be described by a beam with marked and extended side lobes ("dirty beam"), giving rise to a distorted image ("dirty map"). CLEAN removes, by deconvolution, the side lobes without, however, being a true deconvolution method. The result is not a model of the source distribution (to be observable with a perfect very large instrument) but rather a model of what would have been observed with a single dish of the same size as the interferometer ("clean map"). That is, it is a data model and not a source model which would need convolution with a beam to reproduce data. One could also say that the missing Fourier coefficients are interpolated but not extrapolated.

The dirty beam h_{ij} is divided into two parts: The "clean beam" $h_{ij}^{(1)}$ which is essentially the main lobe, or the response of a correspondingly large single dish to a point source, and the sidelobes and main-beam distortions, or "dirt" $h_{ij}^{(2)}$:

$$h_{ij} = h_{ij}^{(1)} + h_{ij}^{(2)} \; . \tag{12}$$

The original image-space data $f_1^{(0)}$ ("dirty map") is deconvolved to a source map $\{m_j\}$ but (in the original version) without additional constraint. The deconvolution result cannot be used directly because it is not smooth enough. Owing to the fact that sidelobes tend to be more extended or at least not less extended than the main lobe, a smooth "image", more or less free of sidelobe effects, can be recovered by convolving the source map with the clean (or "restoring") beam. One actually ends up with an improved (or "restored") data map $f_1^{(1)}$ ("clean map")

$$f_i^{(1)} = \sum_j m_j h_{ij}^{(1)} + r_i = f_i^{(0)} - \sum_j m_j h_{ij}^{(2)} \; . \tag{13}$$

The method is quite ingenious as it avoids such difficult questions as whether or not a smooth image is also a true image. It was also the first method not to neglect missing Fourier coefficients but to choose them according to a reasonably smooth image. Nevertheless, it definitely does not increase the resolution. Therefore, several improvements have been proposed of which we mention only two. First, one can restore with a clean beam that is decreased in size. This is equivalent to including the outer parts of the main lobe into the dirt. Second, one can introduce a constraint such that the deconvolution result is smooth enough to be directly used. This is done in the SSC.

CLEAN has as yet mostly been used in interferometry but at least some of the modern versions are suitable for other problems as well (Becker and Duerbeck 1980).

6 Comparison of methods

There are many deconvolution methods, of which we have mentioned only some, and all have different difficulties. It would certainly be wrong to try to make a linear order of successfulness for the available methods. Even if a method gives a result that looks "good" (meaning that it does not contain obviously improbable or impossible features), it may not be the most reliable one. In general, the best advice as to which methods should be used is to try several ones, and compare the results. Such procedure will quite often provide more information on the probable source distribution than the selection of just one method.

However, some general statements are nevertheless possible. First, no method is hitherto sufficiently understood to know exactly all the advantages and disadvantages, and to know how would be the best interpretation of the results in terms of reliability (as the question whether a slightly extended feature should be interpreted as an unresolved (nearly pointlike) source or as a resolved one). Or to give another example: Inspite of a wealth of theoretical papers on MEM, there is still not even agreement on which form of the entropy should be used. More dangerously, it is not known how much one part of the map may influence the results on other parts of the map.

This is because entropy is a "universal" constraint, not dependent on any details of the measurement. One only knows, from many practical examples, that the mutual influence seems, in most cases, small enough to be neglected.

There is always a competition between smoothness and resolution. The grand design of a map is most easily recognized if the map is very smooth but some essential details may be lost. High resolution tends to overresolve noisy data. The best compromize is probably MEM, with a very smooth image and some superresolution (= resolution beyond that of the data). The claim of some MEM theorists that MEM yields a maximum in possible superresolution is not true. The best resolution so far has been obtained by MIM which, on the other hand, tends to yield a noisier result than other smoothing-constraint methods. The resolution of optimized versions of CLEAN is comparable to that of MEM.

The opinion is widely held (see, *e.g.*, Koch and Anderssen 1987) that the result of a deconvolution should be unique (concave problem). It has been shown that the one-constraint MEM (but not two-constrained versions as that of Reiter and Pfleiderer 1986) as well as the basic CLEAN (Schwarz 1978, Marsh and Richardson 1986) do indeed converge to a unique result, independent of the actual realization of the method in the form of a specific numerical procedure. However, uniqueness is probably quite unimportant. Different methods do give different results, and still we are often unable to choose one as being better than another. The only criterion for the goodness of a solution is whether or not it looks "good" enough in the sense stated above - unless one can compare with better data. However, the most interesting use of deconvolution is, of course, that for the best available data where such comparison is not possible. Non-unique methods do, however, have the disadvantage that the result may depend on the numerical procedure. If different procedures produce different results within one method, one could consider them as varieties of a method and try to find out which variety, if any, works best. The iterational Fourier methods are not unique, the cut-off point of the iterations being empirically determined.

One mandatory feature of uniqueness in constraint methods is that the optimum Lagrange parameter connecting the data fit and the smoothing constraint must be determined by the method itself. It seems to the present writer that this is not necessarily a good approach. Depending on the questions asked, one and the same set of data may be used to emphasize the grand design (large smoothing) or fine details (little smoothing). Some methods therefore allow the choice of the degree of smoothing. A consistent theory of structure (Pfleiderer, in preparation) seems to make the free choice even mandatory.

Unfortunately, all methods which are such simple that they would easily be programmed are also inferior to others. In practice, one therefore has, generally, to rely on what methods an available program library has to offer.

Acknowledgments

Thanks are due to many colleagues for very helpful and clarifying discussions, especially to E.E. Baart, T.J. Cornwell, L. Ericson, S. Gottesman, J.A. Högbom, G.E. Ioup, J. Reiter, U. Schwarz, S. Tan.

References

Ables, J.G., 1974. *Astr. Astrophys. Suppl.*, **15**, 383.

Basistov, Y.A., Goncharski, A.V., Lekht, E.E., Cherepashchuk, A.M., Yagola, A.G., 1979. *Sov. Astr.*, **23**, 248.

Becker, H.J., Duerbeck, H.W., 1980. *Publ. Astr. Soc. Pacific*, **92**, 792.

Biraud, Y., 1969. *Astr. Astrophys.*, **1**, 124.

Bracewell, R.N., Roberts, J.A., 1954. *Austr. J. Phys.*, **7**, 615.

Brault, J.W., White, O.R., 1971. *Astr. Astrophys.*, **13**, 169.

Bryan, R.K., Skilling J., 1980. *Mon. Not. R. astr. Soc.*, **191**, 69.

Cornwell T.J., 1983. *Astr. Astrophys.*, **121**, 281.

Frieden, B.R., 1972. *J. Opt. Soc. Am.*, **62**, 511.

Frieden, B.R., 1975. In *Picture processing and digital filtering (Topics in Appl. Physics No. 6)*, ed. Huang, T.S., Springer, Berlin, p. 177.

Gull, S.F., Daniell, G.J., 1978. *Nature*, **272**, 686.

Helstrom, C.W., 1967. *J. Opt. Soc. Am.*, **57**, 297.

Högbom, J.A., 1974. *Astr. Astrophys. Suppl.*, **15**, 417.

Ioup, G.E., Ioup, J.W., 1983. *Geophys.*, **48**, 1287.

Jansson, P.A., Hunt, R.H., Plyler, E.K., 1970. *J. Opt. Soc. Am.*, **60**, 596.

Jaynes, E.T., 1957. *Phys. Rev.*, **106**, 620.

Jensen, O.G., Ulrych, T., 1973. *Astr.J.*, **78**, 1104.

Jonas, J., 1985. *South Afr. J. Phys.*, **8**, 67.

Koch, I., Anderssen, R.S., 1987. *Astr. Astrophys.*, **183**, 170.

Komesarov, M.M., Narayan, R., Nityanda, R., 1981. *Astr. Astrophys.*, **93**, 269.

Lucy, L.B., 1974. *Astr. J.*, **79**, 745.

Nityananda, R., Narayan, R., 1982. *J. Astrophys. Astr.*, **3**, 419.

Pfleiderer, J., 1985. *South Afr. J. Phys.*, **8**, 73.

Pfleiderer, J., 1988. *Astr. Astrophys.*, **194**, 344.

Pfleiderer, J., Reiter, J., 1982. *Sitz.-Ber. Oest. Akad. Wiss. Math.-Nat. Kl.*, **191**, 591.

Phillips, D.L., 1962. *J. Assoc. Comp. Mach.*, **9**, 84.

Reiter, J., Pfleiderer, J., 1986. *Astr. Astrophys.*, **166**, 381.

Sanromà, M., Estalella, R., 1984. *Astr. Astrophys.*, **133**, 299.

Schwarz, U., 1978. *Astr. Astrophys.*, **65**, 345.

Siska, P.E., 1973. *J. Chem. Phys.*, **59**, 6052.

Subrahmanya, C.R., 1980. *Astr. Astrophys.*, **89**, 132.

Tikhonov, A.N., 1963. *Sov. Math. Dokl.*, **4**, 1035 and 1624.

Tikhonov, A.N., Arsenin, V.Y., 1977. *Solutions of Ill-posed Problems*, Winston & Sons.

Turchin, V.F., Turovtseva, L.S., 1974. *Opt. Spectrosc.*, **36**, 162.

Twomey, S., 1963. *J. Assoc. Comp. Mach.*, **10**, 97.

Twomey, S., 1965. *J. Frank. Inst.*, **279**, 95.

van Cittert, P.H., 1931. *Z. Phys.*, **69**, 298.

Wernecke, S.J., 1977. *Radio Sci.*, **12**, 831.

Wiener, N., 1942. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, MIT Press, Cambridge, MA.

Wittmann, A., 1971. *Solar Phys.*, **21**, 237.