ON THE DANGER OF APPLYING STATISTICAL RECONSTRUCTION METHODS IN THE CASE OF MISSING PHASE INFORMATION

J.C. Dainty, M.A. Fiddy and A.H. Greenaway[+]
Department of Physics, Queen Elizabeth College,
Campden Hill Road, London W8 7AH, U.K.

[+]Département d'Astrophysique, Université de Nice,
Parc Valrose, 06034 Nice Cedex, FRANCE.

I. INTRODUCTION

The use of entropy as a basis for object/image reconstruction procedures is not new[1], but with the appearance of new, faster algorithms[2] the actual use of these algorithms for the reconstruction of objects from 'real' data is likely to increase.

The purpose of this contribution is not to discourage such applications, but to illustrate that, under certain circumstances, there is a need for caution in interpreting the results obtained from such algorithms. Specifically, we shall show that the application of statistical methods to problems of object reconstruction, in situations where only the modulus of the object Fourier transform is known, could lead to wholly false conclusions. Indeed, we shall primarily be concerned here with situations for which *there is no 'correct' solution*. In such situations it is pointless to speak of 'safe' object reconstruction algorithms. The important point here is that the user of a statistically based 'object reconstruction algorithm' may be totally ignorant of whether or not he is working in this régime.

The problem thus formulated, i.e. the reconstruction of an object from its autocorrelation, is the so-called phase problem, which occurs in fields as diverse as electron microscopy, radio and optical astronomy and X-ray diffraction. It has long been recognised that, in general, the solution to such problems cannot be unique. In all cases it is necessary to provide additional information, such as the phase of the object Fourier transform, in order to resolve the inherent ambiguities associated with this problem.

The key point to this paper, is that to speak of 'safe' reconstruction procedures, one must show that the problem posed has a unique solution. If one can show that this is indeed the case, any algorithm which yields a solution consistent with the defined problem will give the correct solution. In this sense all such algorithms may be regarded as 'safe'.

C. van Schooneveld (ed.), Image Formation from Coherence Functions in Astronomy, 95-101.

https://doi.org/10.1017/S0252921100074789 Published online by Cambridge University Press

The choice of algorithm will then be dictated by the speed and ease with which it may be implemented, and its stability against the effects of noise and data truncation. It is in this latter sense that maximum entropy algorithms may be regarded as 'safe'.

By application of the theory of entire functions we shall demonstrate the fundamental ambiguity associated with phase problems and illustrate this with some specific examples. We shall also give consideration to the conditions under which the degree of ambiguity may be reduced, or where a unique solution may be obtained.

Initially, we shall consider the situation where the recorded modulus data is perfect and complete. The situation where the data is noisy and/or where some phase information is available, will be considered later.

2. THEORY

It is not our intention here to give a rigourous treatment of the phase problem in terms of entire function theory, but we shall briefly indicate the results of this analysis[3].

Let the object field to be reconstructed be represented by $f(\xi)$ and its Fourier transform by $F(x)$, then

$$F(x) = \int_a^b d\xi \; f(\xi) \; \exp(-i\xi x) \tag{1}$$

where the limits of integration (a,b) will always be finite, either because of the finite field of view of the instrument or due to the finite extent of the object. Attention will be confined to this one-dimensional case, since this simplifies the analysis and illustrates all the features we wish to discuss.

The use of equation (1) to describe the system, imposes some severe and far reaching restrictions on the behaviour of $F(x)$. Specifically, any $F(x)$ which is a solution to an equation such as (1), must be the limit on the real(x-) axis of an analytic function which is of order unity and of finite type. Thus $F(z)$, which may be generated by replacing x with $z = x + iy$ in equation (1), is analytic for all finite values of z and may be specified, to within trivial ambiguities, from a knowledge of the set of points at which it is identically zero. In this respect $F(z)$ is like a polynomial whose roots (i.e. zeros) must occur at isolated points. For an $F(z)$ generated from equation (1) there will always be a denumerable infinity of such roots. These zeros are distributed so that in any finite region of the z-plane there will only be a finite number of roots, their density being proportional to $(b-a)$[4]. Furthermore, all such zeros must lie close to the real axis and will tend to lie along the x-axis at the Nyquist sampling rate as $x \to \pm\infty$. Thus with any finite section of the x-axis we may associate a finite number of zeros. It has been noted[5] that

https://doi.org/10.1017/S0252921100074789 Published online by Cambridge University Press

this is the same as the number of degrees of freedom of the system.

Formally, we may express the relationship between F(x) and the roots of F(z) by

$$F(x) = F(0) \exp(-ix\{\frac{b-a}{2}\}) \prod_{j=1}^{\infty} (1-\frac{x}{z_j}) \tag{2}$$

where z_j, j=1,2,... are the roots of F(z) arranged such that $|z_{j+1}| \geq |z_j|$. In writing equation (2) it has been assumed that $F(0) \neq 0$; this is always true in astronomy where objects are real and positive; the condition can be relaxed if required.

When the phase of F(x) has been lost, the observable magnitude is the scattered intensity, which is proportional to

$$I(x) = |F(x)|^2 = F(x)\, F^{\star}(x) \tag{3}$$

Now visual inspection of (2) shows that $F^{\star}(x)$ will have its roots at $z_j^{\star}$, j=1,2,... . Thus I(z), the continuation of I(x) into the z-plane has zeros at z_j and $z_j^{\star}$ and so has twice as many zeros as F(x). Thus, when attempting to reconstruct F(x) , or equivalently f(ξ), from I(x) roots of F(x) which occur on the x-axis may be identified with certainty but for the complex roots, i.e. those not on the real axis, it is impossible to decide whether F(z) has a root above or below the x-axis. If there are M complex zeros, the repetition of this twofold ambiguity for each such zero given rise to 2^M possible alternative solutions for F(x), all of which correspond to the same intensity, I(x), and all of which have inverse Fourier transforms which are confined to $\xi \varepsilon (a,b)$. *On the basis of measurements of* I(x) *alone, there is no possible way to distinguish between these alternatives.* It has been known for some time that, with perfect data and given the constraints imposed by equation (1), this is the *only* source of ambiguity.[6]

In some situations it is possible to resolve or reduce this ambiguity by making f(ξ) satisfy certain *a priori* conditions. For example, if f(ξ) is forced to take a known value over some ξ-range within the interval (a,b)[7], or if it is known *a priori* that all zeros lie in one half of the complex plane[3], a unique solution may be ensured. A special case of all zeros being in one half plane occurs when they are all real. The existence of a unique solution in this case has been exploited to determine the structure of nerve myelin from X-ray data[8]. If f(ξ) is known to be real, the zeros of F(z) must satisfy certain symmetry properties which reduce the ambiguity by a factor of $2^{M/2}$.

Given that difficulties of this kind exist with perfect intensity data it is essential that any reconstruction algorithm which uses intensity data is capable of resolving this implicit ambiguity. It is not clear that statistically based 'object reconstruction' algorithms, such as maximum entropy, can take any account of such ambiguities, the importance of which are illustrated in the following section.

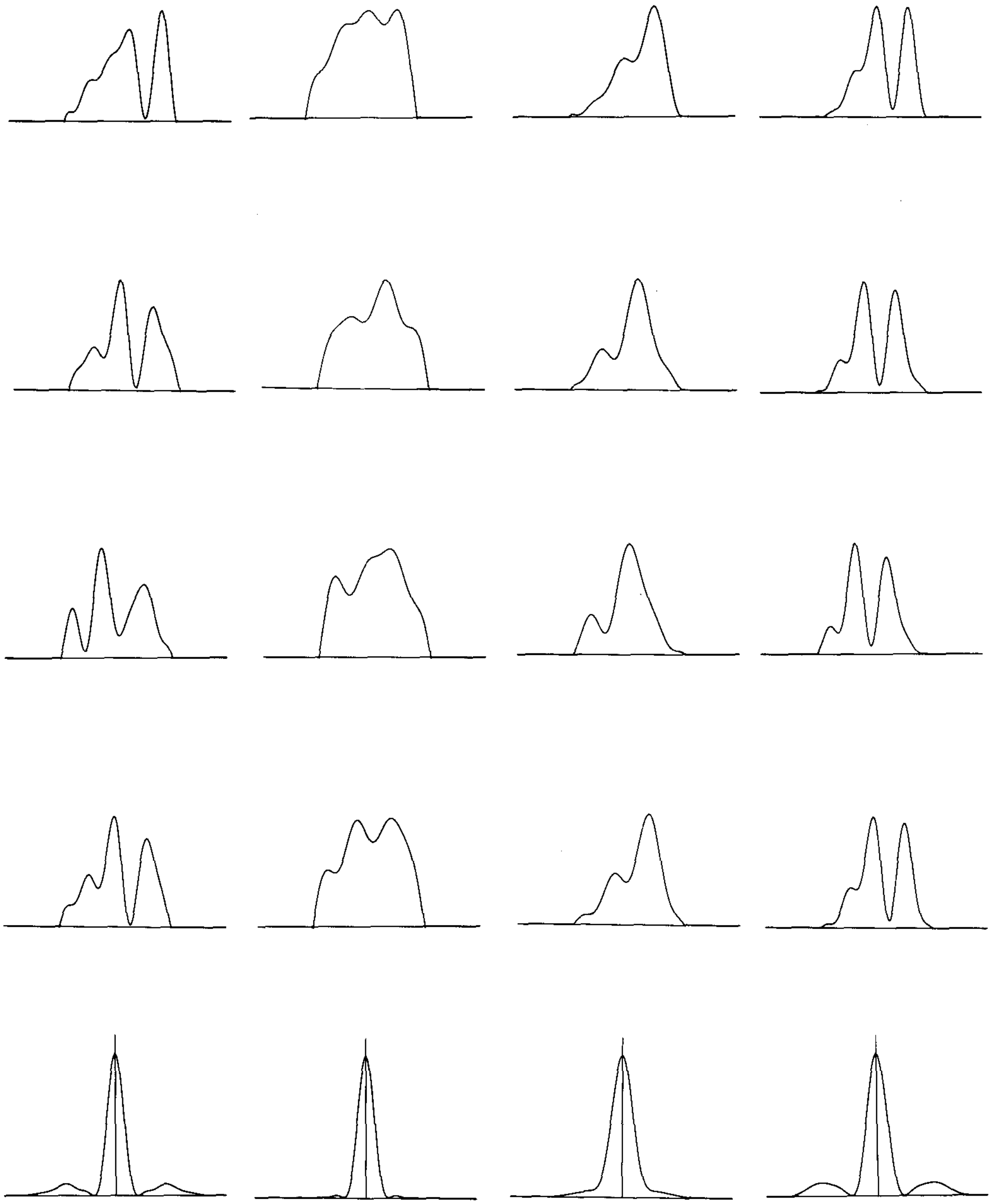

Figure 1 Four examples of ambiguous object reconstruction. The upper four graphs in each column show four different objects having the same I(x), which is shown in the lower graph. For discussion see the text.

3. EXAMPLES OF AMBIGUOUS RECONSTRUCTIONS

In figure 1 we show four examples of ambiguity in object reconstruction from measurements of I(x) alone. In each example we have constructed an F(z) which has just six complex zeros, the remainder being real. Each example therefore has 2^6 possible solutions for $f(\xi)$, all wholly consistent with equation (1). If we specify that $f(\xi)$ must be real, then there must be a symmetry of the zeros of F(z) about the y-axis. Enforcing this symmetry reduces the ambiguity to 2^3 equally acceptable real solutions for each example. Now, reflecting any given distribution of zeros about the x-axis is equivalent to replacing $f(\xi)$ by $f(-\xi)$. This difference we shall neglect, so reducing the ambiguity by another factor of two. For each example this leaves four equally valid, real and positive solutions for $f(\xi)$. These are shown in figure 1.

Each example occupies one column of the figure, the upper four graphs of each column being the four possible object reconstructions which may be made from the observed intensity shown in the lower graph of each column. In each case the top graph of each column corresponds to the minimal or Hilbert phase solution[3], with all zeros in one half plane.

Note that for none of these examples may one speak of a 'correct solution', in each case the four alternatives represent equally valid object reconstructions. There can be no question of *any* algorithm providing the 'correct solution' unless other information is available. The examples in figure 1 show that this statement is non-trivial.

In the first example, shown in the left hand column, one cannot even say with certainty how many peaks there are in the object distribution. In the second example it is clear that the object consists of a single peak of reasonably well defined width, but what can one say with certainty about the structure superposed on that peak? The ambiguities in the third and fourth examples are less severe but even here the width, relative position and height of each peak is somewhat uncertain.

4. INCOMPLETE AND NOISY DATA, WITH AND WITHOUT PHASES

So far it has been assumed that no information about the phase of F(x) is available. Clearly, if both the modulus and phase of F(x) are available and the data is perfect and complete, the ambiguity is resolved, thus the zero positions are specified exactly.

If we incorporate the effects of noise and data truncation into our previous description, it is evident that, even when phase information is available, we can only obtain *estimates* for the positions of the zeros of F(z). If no phase information is available and I(x) is contaminated with errors, not only is the exact position of each zero uncertain but the fundamental ambiguity, i.e. determining in which half plane each zero belongs, must also be resolved. Since errors are inherent in measured data we must resign ourselves to defining regions in which the zeros must be located, rather than exact zero positions. More exactly, we should define a probability density function for the position of each zero. *In the absence of phase information this p.d.f. will be symmetric with respect to the real axis.*

When some phase information is available we may expect the p.d.f.s to exhibit a maximum to one side of the real axis. Under these circumstances some algorithm for selecting an, in some sense, optimum set of zero positions consistent with the p.d.f.'s would be useful. Such a procedure, based on the theory of entire functions, has already been proposed and used to reduce radio astronomical data.[9] This method relies on the experimental observation that when both the phase and modulus are measured experimentally the uncertainty in the phase frequently exceeds that in the modulus. One may therefore use I(x) to determine the zero co-ordinates, with the phase used merely to decide in which half plane each zero belongs. The resulting reconstructions may be considerably more accurate than those made using the raw data. The maximum entropy algorithm may well fulfil the same rôle without the need to invoke the theory of entire functions explicitly.

5. CONCLUSIONS

We have demonstrated the well known result that in the absence of additional information it is, generally, impossible to restore an object unambiguously from measurements of the modulus of its Fourier transform. We have shown with counter examples that positivity of the solution does not constitute the additional information needed to resolve this ambiguity. Nor do statistically based algorithms, e.g. maximum entropy, provide such information. Thus they offer no way out of the dilemma and they cannot be regarded as giving a 'safe', i.e. unique, restoration of the objects from the modulus of their Fourier transforms. Arguments that errors in real data may negate these conclusions are equivalent to suggesting that ignorance provides a suitable basis for inference.

If phase information is available, the maximum entropy algorithm may represent a useful approach to the problem of selecting an optimum set of zero positions consistent with imperfect data. In the absence of phase information, statistical algorithms should provide an optimized version of one of the 2^M possible object reconstructions (perhaps the Hilbert solution). They will not, in general, provide any information about the other possible solutions. This 'lost' information could be retrieved by using the optimized $f(\xi)$ to calculate the roots of $F(z)$ and proceeding as above. Note, however, that in the examples discussed here M has been deliberately kept small. In 'real' experiments it may be much larger with a consequent increase in the number and range of solutions.

ACKNOWLEDGEMENTS

This work has been partially supported by the S.R.C. in the U.K. and by the C.N.R.S. in France under the European Scientific Exchange Scheme.

REFERENCES

1 Frieden B.R. : 1972,*J.Opt.Soc.Am.*,62,pp511-518
2 Gull S.F. and Daniell G.J. : 1978,*Nature*,272,pp686-690
3 Burge R.E., Fiddy M.A., Greenaway A.H. and Ross G. : 1976, *Proc.Roy.Soc.*,A350,pp191-212.
4 Titchmarsh E.C. : 1926,*Proc.Lond.Math.Soc.*,25,pp283-302.
5 Burge R.E., Fiddy M.A., Greenaway A.H. and Ross G. : 1976 *Recent Advances In Optical Physics*,(ICO-10, Prague),pp 687-694
6 Walther A. : 1963,*Optica Acta*,10,pp41-49.
7 Greenaway A.H. : 1977,*Optics Lett.*,pp10-12.
8 Burge R.E., Fiddy M.A. and Wheeler M.W.L. : submitted to *J.Mol.Biol.*
9 Bates R.H.T., and Napier J. : 1972,*Mon.Not.R.astr.Soc.*,158,pp405-424.

DISCUSSION

Comment J.E. BALDWIN
In the complex plane analysis it is easy, given one solution, to produce any other by transferring zeros across the real axis. Starting with a solution comprising point sources separated by regions of zero intensity, I believe that all the other solutions have strong negative features, so that the solution for a point source case is almost certainly unique.

For extended objects it may be that regions of zero intensity inside the area within which the source lies, even though we do not know where they are, may be sufficient to make the solution unique. Have you investigated this possibility or think it likely to be useful?

Reply A.H. GREENAWAY
Although I have not experimented with distributions comprised of isolated point sources I am inclined to agree with you that, if we ignore 180^{o} ambiguities, the solution is probably rendered unique by application of the positivity constraint.

With regard to extended objects, the exact knowledge of any small internal region of the object is certainly sufficient to allow a unique reconstruction from the modulus data (see ref. 7), even in the absence of a positivity constraint. Your suggestion that for objects with internal gaps the positivity constraint alone may be sufficient is very interesting. One may certainly quote counter examples for which this information does not lead to a unique solution - for example an object consisting of two identical regions separated by a blank space would not give a unique solution if the regions in question individually give rise to ambiguities of the type shown in figure 1. If the two regions are dissimilar, but have one or more complex zeros in common one would also expect ambiguities which would not be resolved by these constraints. However, these are rather trivial examples and it may well be that the use of positivity with an internal gap in the object will, in many cases, lead to a unique solution.

https://doi.org/10.1017/S0252921100074789 Published online by Cambridge University Press