THE INTRODUCTION OF A PRIORI KNOWLEDGE IN CERTAIN PROCESSING ALGORITHMS

J.A. Högbom
Stockholm Observatory

A common problem in radio synthesis work is that of determining the brightness at N grid points in the map domain when there are only $n < N$ independent interferometer measurements available. The missing $(N-n)$ equations can in principle be replaced by an equivalent amount of information in the form of a priori knowledge about the brightness distribution. One way of doing this is to add an equation of the form

$$F(\text{map}) = \max , \qquad (1)$$

i.e. some function of the brightness distribution is maximized subject to the condition that the map be compatible with the n measurements. This automatically gives $(N-n)$ new equations, leaving us in the pleasant situation of having as many equations as there are unknowns.

The expression (1) is very general and potentially able to absorb all kinds of a priori knowledge that we care to introduce. On the other hand, this generality is inconvenient when designing efficient processing algorithms. Several algorithms in current use are instead based on a simplified version of (1):

$$\Sigma f(x_i) = \max , \qquad (2)$$

the sum to be taken over all the values x_i of the brightness at the N grid points of the map. The function $f(x)$ is to be specified in such a way that it becomes an expression of the a priori information. The left hand side of (2) is a function of the individual values x_i but not of their spatial distribution over the map. Hence, this simple formulation cannot be used to express a priori information about spatial relations between features on the map, but it does permit the use of information related to the statistics of the values x_i.

The argument has been advanced that the function $f(x)$ should relate to what may be termed the 'entropy' of the map (Ables 1974, Gull and Daniell 1978). The resulting maximum entropy map becomes, in a certain sense, that which contains the least amount of information -

C. van Schooneveld (ed.), Image Formation from Coherence Functions in Astronomy, 237–239.

and hence presumably the least amount of false information - among all possible maps that are compatible with the measurements. It has even been claimed at this symposium that, of all possible inversion procedures, the maximum entropy method is fundamental and unique in some deep philosophical sense.

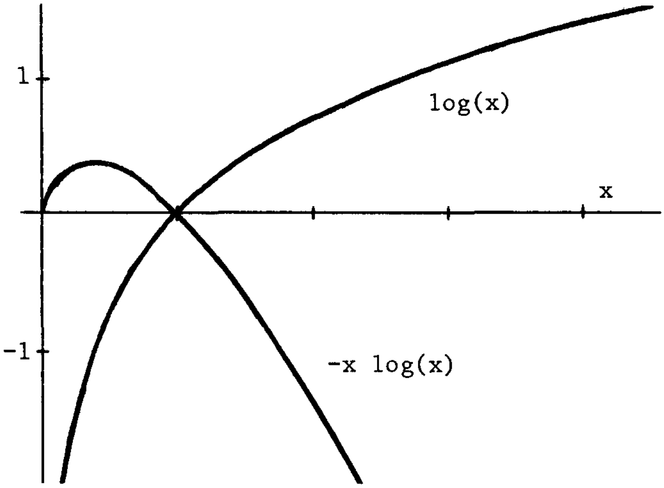
There are two main competing views as to what should be the form of the function f(x) in this context. They are:

$$f(x) = \log(x), \quad \text{and} \tag{3}$$
$$f(x) = -x \log(x) \tag{4}$$

At first sight these two functions appear to be very different (Figure 1). If one is fundamental and unique, then one would expect the other to perform very badly in comparison. However, when confronted with real data, both alternatives do quite well.

Figure 1

The functions log(x) and -x log(x)



A good expression for the entropy of a radio brightness distribution - if such a thing exists - is not necessarily identical to one that applies to idealized distributions encountered in physics or mathematics. However, there may well be similarities. Thus one should concentrate on the similarities between (3) and (4) in order to find out what makes these functions work, rather than debate which of the two is the one and only expression for the map entropy.

Both functions are real for $x > 0$ and their first derivatives go to infinity as x approaches zero. This represents a convenient way of introducing the positive constraint, the a priori knowledge that brightness distributions are exclusively positive. The progressive steepening as x approaches zero means that eq(2) will prevent map amplitudes x_i from becoming zero or negative.

https://doi.org/10.1017/S0252921100074935 Published online by Cambridge University Press

The two functions are convex, their second derivatives are negative everywhere. This has the effect of biasing the map in favour of a uniform distribution. Any unnecessary dispersion of the amplitudes x_i about their mean value (assuming this to be specified in advance) does not satisfy eq(2) since it will lower the entropy of the map. Gull has pointed out that the form $f(x) = -x \log(x)$ will give maps over which the relative amplitude errors tend to be constant. This is a desirable property when one wants to find out what radio features are present in the mapped region of sky. It is less desirable if one wants to derive accurate structures and intensities of those features. Special precautions are needed in order to avoid overinterpretation of noisy data in low brightness areas.

As noted above, eq(2) cannot be used to introduce a priori knowledge about the spatial distribution of the values x_i over the map. Such information must be provided by the n measurements. In interferometry these specify certain spatial frequency components that are present in the brightness distribution. Some dispersion of the x_i values is required to accomodate these components but, since any unnecessary dispersion is discouraged by eq(2), large amplitudes will occur where they most effectively contribute to satisfying the measurements, i.e. where sources are likely to be situated.

The question of what is the 'best' map that is compatible with the measurements and the a priori knowledge has different answers depending on what one wants to do with the map. The maximum entropy approach is very valuable because it has demonstrated the potentialities of the simple approach represented by eq(2). The 'entropy' functions suggested so far may be regarded as examples of a class of functions $f(x)$ which allows the user some freedom in matching the general properties of the map to the type of information he wants to extract from his measurements.

REFERENCES

Ables, J.G.: 1974, "Astron. Astrophys. Suppl." 15, p. 383.
Gull, S.F., Daniell, G.J.: 1978, "Nature" 272, p. 686.

https://doi.org/10.1017/S0252921100074935 Published online by Cambridge University Press