INTERFEROMETER GROUP INTERNAL REPORT

MAXIMUM ENTROPY SPECTRAL ANALYSIS

by

Louis J. Gross

THE CONCEPT

The idea upon which maximum entropy analysis is based is essentially a very simple one. We merely aim to perform a transformation upon our data in such a manner that we have introduced the least amount of bias in the results, while at the same time making full use of what information we do have. The concept is explained in a very readable and erudite manner by John Ables (ref. 1), and I shall not attempt to improve upon it, though I will set out some of the basics here.

The entire method hinges upon performing the transformation in such a way that minimizes the "information" that we have added to our data. We wish to choose, from all the possible spectra whose transform will be our data, that spectrum which has the highest entropy (ignorance), i.e. the one to which we have added the least amount of "information". To accomplish this, we must use some expression to represent the entropy of the spectrum. Bartlett (ref. 2) defines the entropy of a continuous stochastic process to be

$$H = -\int_{-\infty}^{\infty} f(x) \log f(x) \, dx$$

where f is the probability density function of the signal amplitude. We can think of any signal with spectrum $S(\nu)$ as simply white noise passed through a linear filter whose gain is $S(\nu)$. Now the difference between the entropies of the input and output signals from this filter is shown by Bartlett to be a function of $S(\nu)$ alone and is in fact given by

$$\Delta H = \int_{-\infty}^{\infty} \log S(\nu) \, d\nu$$

Since we want our spectrum to have the greatest ignorance, we wish to make the entropy gain as great as possible. So we wish to maximize ΔH.

Therefore, we can now state our problem. Given a visibility function sampled at 2N+1 points $X_0 = \{\rho_n\}$, $n = 0, \pm 1, \ldots, \pm N$, we wish to choose that spectrum $S_0(\nu)$ such that

$$I = \int_{-\infty}^{\infty} \log S_0(\nu)\, d\nu$$

is a maximum under the constraints

$$\rho_n = \int_{-\infty}^{\infty} S_0(\nu)\, e^{2\pi j n \nu}\, d\nu \qquad \text{for } n = 0, \pm 1, \ldots, \pm N$$

A SOLUTION

The above problem can be solved analytically using a constrained calculus of variations technique. We first introduce Lagrangian multipliers λ_n, $n = 0, \pm 1, \ldots, \pm N$ for each of our constraints. Then the functional we wish to maximize is

$$J = \int_{-\infty}^{\infty} \log S(\nu)\, d\nu - \sum_{n=-N}^{N} \lambda_n \left[\int_{-\infty}^{\infty} S(\nu)\, e^{2\pi j n \nu}\, d\nu - \rho_n \right]$$

Now, taking the first variation of J, we first form

$$J(\epsilon) = \int_{-\infty}^{\infty} \log\left[S(\nu) + \epsilon\, \eta(\nu) \right] d\nu - \sum_{n=-N}^{N} \lambda_n \left[\int_{-\infty}^{\infty} \left[S(\nu) + \epsilon\, \eta(\nu) \right] e^{2\pi j n \nu}\, d\nu - \rho_n \right]$$

where ϵ is an arbitrary constant and $\eta(\nu)$ is an arbitrary function. The first variation is then $J'(\epsilon)$, the derivative of J with respect to ϵ, and a necessary condition for an extremum is that $J'(0) = 0$. So we have

$$J'(\epsilon) = \int_{-\infty}^{\infty} \frac{\eta(\nu)}{S(\nu) + \epsilon\, \eta(\nu)}\, d\nu - \sum_{n=-N}^{N} \lambda_n \int_{-\infty}^{\infty} \eta(\nu)\, e^{2\pi j n \nu}\, d\nu$$

And we must have,

$$J'(0) = \int_{-\infty}^{\infty} \frac{\eta(\nu)}{S(\nu)}\, d\nu - \int_{-\infty}^{\infty} \sum_{n=-N}^{N} \lambda_n\, \eta(\nu)\, e^{2\pi j n \nu}\, d\nu$$

$$= \int_{-\infty}^{\infty} \eta(\nu) \left[\frac{1}{S(\nu)} - \sum_{n=-N}^{N} \lambda_n e^{2\pi j n \nu} \right] d\nu = 0$$

Since $\eta(\nu)$ is any arbitrary function, we have

$$S(\nu) = \frac{1}{\sum_{n=-N}^{N} \lambda_n e^{2\pi j n \nu}}$$

To ensure that this maximizes J, a necessary condition is that $J''(0) \leq 0$ and since

$$J''(\epsilon) = \int_{-\infty}^{\infty} \left(- \frac{\eta^2(\nu)}{[S(\nu) + \epsilon\, \eta(\nu)]^2} \right) d\nu$$

$$J''(0) = - \int_{-\infty}^{\infty} \frac{\eta^2(\nu)}{S^2(\nu)}\, d\nu < 0$$

the above condition on $S(\nu)$ will indeed maximize J.

We now must solve for the λ_n's. To do this, we set up 2N+1 simultaneous linear equations by looking at

$$\sum_{n=-N}^{N} \rho_{n+r}\, \lambda_n \,, \quad \text{for } r = 0, -1, -2, \dots, -2N$$

For $r = 0$, we have

$$\sum_{n=-N}^{N} \rho_n \lambda_n = \sum_{n=-N}^{N} \lambda_n \int S(\nu)\, e^{2\pi j n \nu}\, d\nu$$

$$= \sum_{n=-N}^{N} \lambda_n \int \frac{e^{2\pi j n \nu}}{\sum_{k=-N}^{N} \lambda_k e^{2\pi j k \nu}}\, d\nu$$

$$= \int \frac{\sum_{n=-N}^{N} \lambda_n e^{2\pi j n \nu}}{\sum_{k=-N}^{N} \lambda_k e^{2\pi k j \nu}}\, d\nu = \int_{-\nu_N}^{\nu_N} d\nu = 2\nu_N$$

where we integrate over the Nyquist interval, $\nu_N = 1/2\Delta\tau$. In general we take $\Delta\tau = 1$, so

$$\sum_{n=-N}^{N} \lambda_n \rho_n = 1$$

Allowing r to take on its other values, we have

$$\rho_{n+r} = \int S(\nu)\, e^{2\pi j \nu (n+r)}\, d\nu = \int \frac{e^{2\pi j n \nu}\, e^{2\pi j r \nu}}{\sum_{k=-N}^{N} \lambda_k e^{2\pi j k \nu}}\, d\nu$$

And,

$$\sum_{n=-N}^{N} \rho_{n+r} \lambda_n = \sum_{n=-N}^{N} \lambda_n \int \frac{e^{2\pi j n \nu}\, e^{2\pi j r \nu}}{\sum_{k=-N}^{N} \lambda_k e^{2\pi j k \nu}}\, d\nu$$

$$= \int_{-1/2}^{1/2} e^{2\pi j \nu r}\, d\nu$$

$$= \int_{-1/2}^{1/2} \left[\cos(2\pi \nu r) + j \sin(2\pi \nu r) \right] d\nu = 0$$

Due to the periodicity of the discrete fourier transform, we have $\rho_n = \rho_{n \pm (2N+1)}$ for all n . The above linear equations can be expressed in matrix form as:

$$\begin{pmatrix} \rho_{-N} & \rho_{-N+1} & \cdots & \rho_N \\ \rho_N & \rho_{-N} & \cdots & \rho_{N-1} \\ \vdots & \vdots & & \vdots \\ \rho_{-N+1} & \rho_{-N+2} & \cdots & \rho_{-N} \end{pmatrix} \begin{pmatrix} \lambda_{-N} \\ \lambda_{-N+1} \\ \vdots \\ \lambda_N \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

To solve for the λ_n's , we then invert the coefficient matrix and the will be the first column of this inverted matrix. Then our spectrum is simply given by

$$S(\nu) = \frac{1}{\sum_{n=-N}^{N} \lambda_n e^{2\pi j n \nu}}$$

Ables goes one step further than this. The above coefficient matrix is of a special type, called circulant, since each row is the above row shifted one element to the right. Through a technique unfamiliar to me, this circulant matrix can be reduced to a simpler N+1 by N+1 Toeplitz matrix, for which there exist very efficient inversion methods. In my work I have not used this technique, but it is definitely a more efficient method and should be used for any large scale use of one dimensional maximum entropy techniques.There may indeed be other useful ways to solve the above set of equations, for example using the properties of the eigenvalues of Toeplitz and circulant matrixes (see ref. 3).

RESULTS AND EXTENSIONS

Figure 1 shows the data for a one dimensional strip scan of NGC7027 taken on the interferometer at XBAND by Drs. Balick and Bignell. The data is equally spaced every 300 meters from 300 to 2700 and I have filled the center cell with the maximum amplitude in the data. Figure 2 shows a discrete fourier transform of this data applying unit weight to each spacing, therefore giving the best resolution. Figure 3 shows a Maximum entropy spectrum of the same data.

As you can see, maximum entropy seems to have resolved one of the components into a double. In maximum entropy spectra of other nearby strips of constant hour angle, this double structure within a component persists, so it is not just a quirk in that particular strip scan. In a full synthesis, after cleaning, the source has only 2 components. As to which method more accurately represents what is really up there, I shall leave that to your judgement.

In performing MEM analysis, it is necessary to assume a value for the center cell, or total flux of the source. By assuming some value for this we are violating Ables's "First Principle of Data Reduction", but not in too harsh a manner. At any rate, I feel that using the maximum amplitude in your data as an estimate of the total flux is certainly preferable to the value zero which is most frequently used.

Since MEM analysis deals with the log of the spectrum, it is necessary to keep the spectrum positive. This is done by adding a constant impulse to the center cell, thereby causing a constant DC offset in the spectrum, which can be subtracted at the end of the analysis. I have found that the amount of offset you add does have some effect on the spectrum. The basic features remain the same, along with their relative amplitudes, however the resolution between features seems to change. I have been unable so far to come up with any reason for this effect. In general, I have found that a center offset of approximately 5 times the maximum amplitude seems to work best.

At present I know of no qualitative method to do a two dimensional maximum entropy analysis. It certainly would be possible to do a large number of one dimensional strip scans at varying hour angles, and then combine them to form a two dimensional map. However, this method would be rather laborious and time consuming. There seems little practical use in deriving a MEM analysis for equally spaced two dimensional data due to the ellipticity of the U-V plane tracks. As for non-equally spaced one dimensional data, I think that most of the above variational theory holds with the minor changes caused by the loss of a certain number of constraints. It is the solution for the Lagrangian multipliers that presents the main problem. All in all, I feel that MEM analysis may be very useful in interferometry, however it will probably be much more time consuming than the current FFT methods.

<u>MENTROPY</u> - A Maximum Entropy Analysis Program For Equally Spaced One Dimensional Interferometer Data

MENTROPY is a FORTRAN program which performs a MEM analysis on equally spaced data taken along a constant hour angle line in the U-V plane. The analysis is done in the manner described above. The program can handle up to 15 points on one side of the hour angle line and assumes the plane is hermitian. The value assumed for the total flux (center cell) is the maximum amplitude of the input points. The DC offset added to the center cell is 4 times this maximum amplitude, and this offset is subtracted from the transformed data at the end of the analysis. The transformed output array can be chosen to be of any length up to 131. Its length should be odd and a recommended value is approximately 8 times the number of data points used.

MENTROPY uses one card input file. The first card contains the desired field of view in seconds of arc (FLD), then the length of the output array (ISIZE) in (F10.0,I5) format. If 1. is used for the field of view, the field will be the full Nyquist interval.If a field of view greater than the Nyquist interval (given by 206265./Δ W, where Δ W is the sampling interval in wavelengths) is used, the spectrum will repeat itself. The first 2 input data points are used to determine Δ W. If ISIZE is left blank, the default value is 73.

Following this card are up to 15 cards with (Amplitude, Phase in degrees, U in wavelengths, V in wavelengths) in (4F8.0) format. This is designed to use as input the cards outputted by INTQMAP specifying PARM=Punch.

The output is:

(1) List of the input data

(2) A plot of the real and imaginary parts of the input data.

(3) The upper left hand 3x3 corner of the coefficient matrix times its calculated inverse. This should be close to an identity matrix, and is meant to be used as a test of the inversion(done by Gauss-Jordan method). It is possible that with certain data the matrix will be highly singular.

(4) The real part of the MEM spectrum.

The program takes about 40 seconds CPU to do one spectrum. See Bill Meredith for a copy of the deck.

ACKNOWLEDGEMENTS

I would like to thank Dr. John Ables for first explaining MEM analysis to me, and Drs. Bruce Balick and Peter Napier for their continuous encouragement and assistance.

REFERENCES

(1) Ables, J. - "Maximum Entropy Spectral Analysis",
Paper presented at Symposium on Collection and Analysis of Astronomical Data , Charlottesville, Nov. 1972

(2) Bartlett, M. S. - "An Introduction to Stochastic Processes" (2nd ed.)
Cambridge University Press, 1966

(3) Gray, R. M. - " On the Asymptotic Eigenvalue Distribution of Toeplitz Matrixes", IEEE Trans. Inf. Theory, Nov. 1972

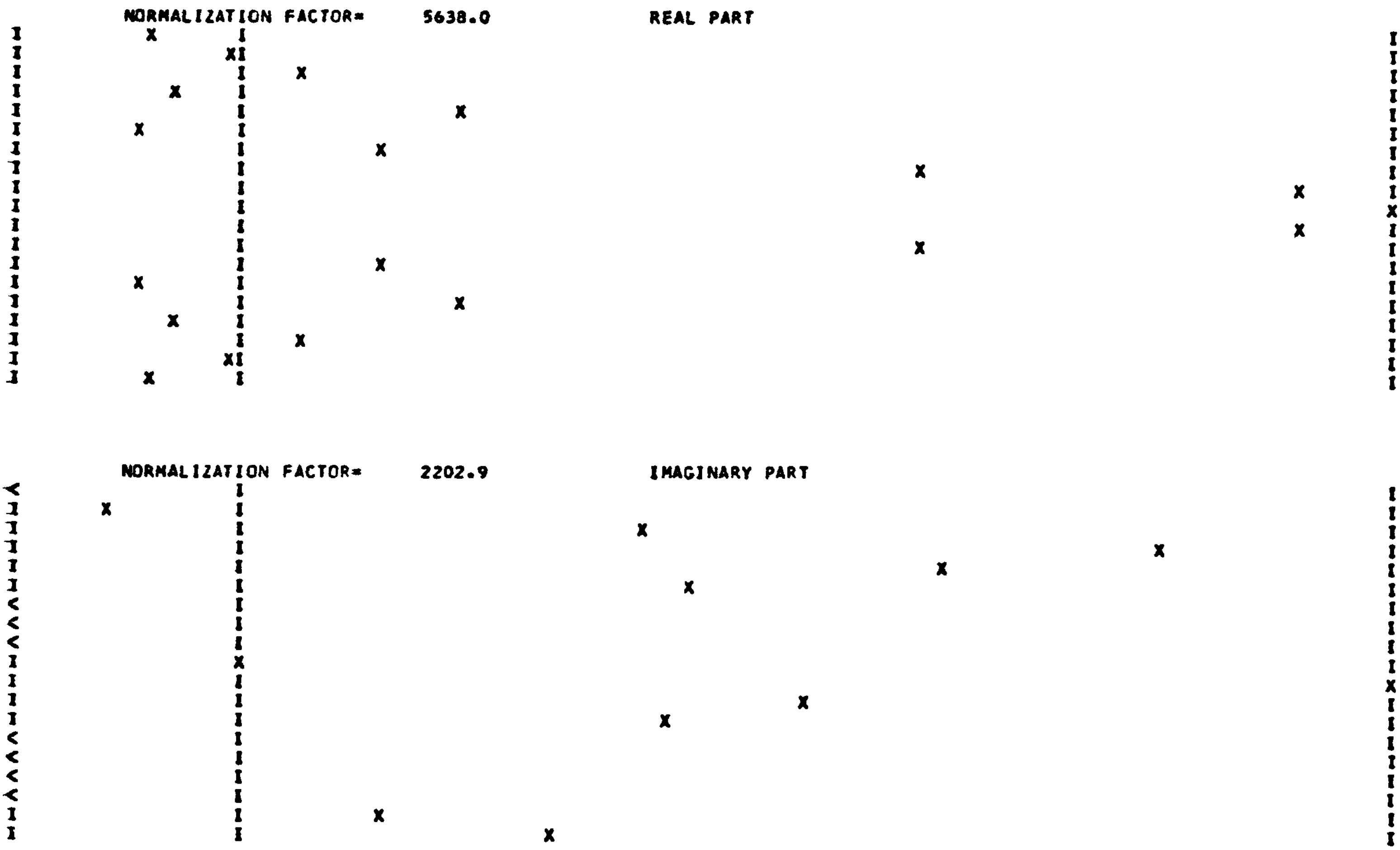
Below is a list of various references dealing with MEManalysis.

Frieden, B. R. - "Restoring with Maximum Likelihood and Maximum Entropy",
JOSA, April,1972

Frieden, B. R. and Burke, J. J. - "Restoring with Maximum Entropy II",
JOSA, Oct. 1972

van den Bos, A. - "Alternative Interpretation of Maximum Entropy Spectal Analysis", IEEE Trans. Inf. Theory, July, 1971

Edward, J. A. and Fitelson, M. M. - "Notes on Maximum Entropy Processing",
IEEE Trans. Inf. Theory, March 1973



NGC 7027 XBAND Position Angle 44°

Figure 1

NORMALIZATION FACTOR= 27558.

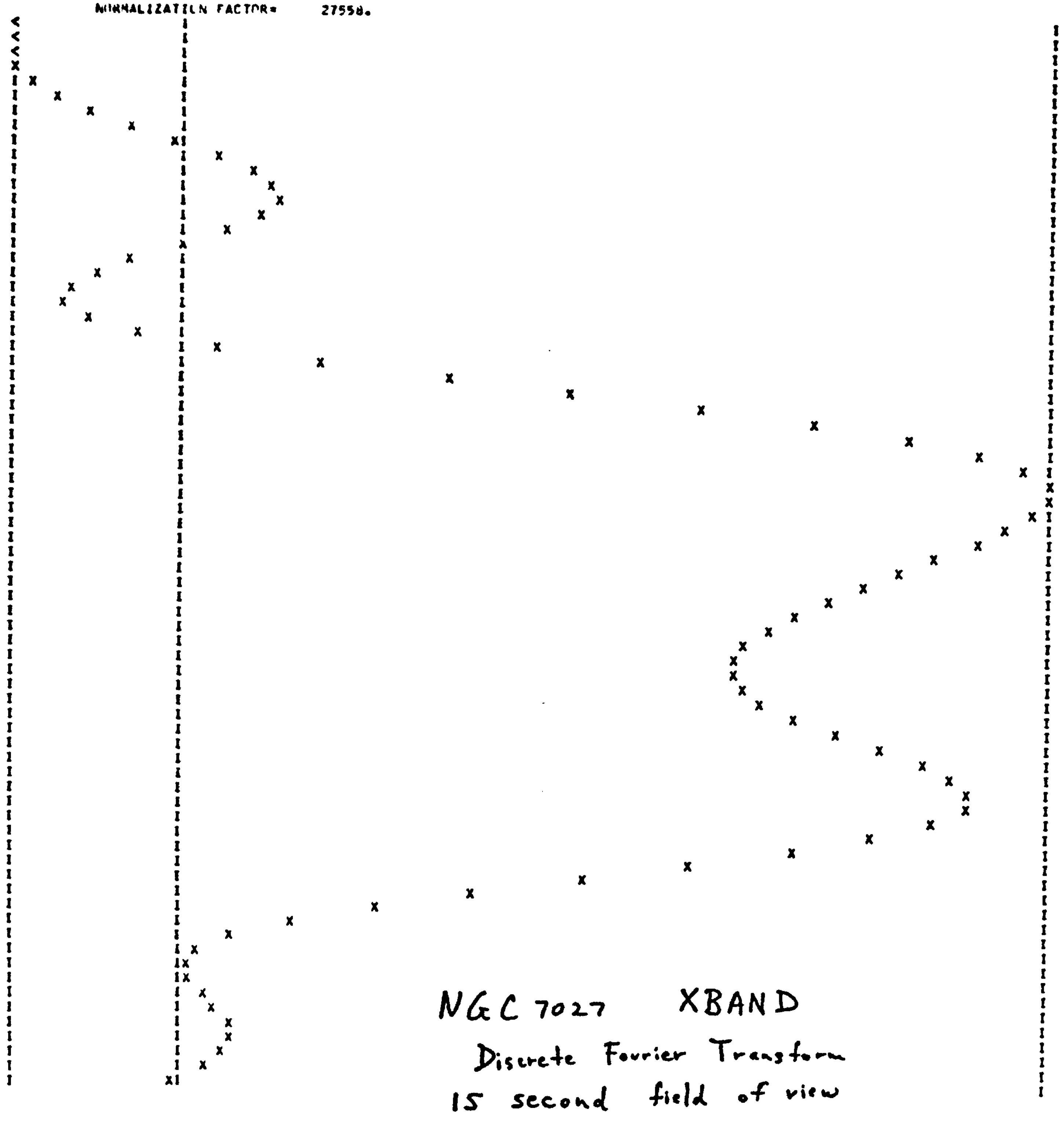
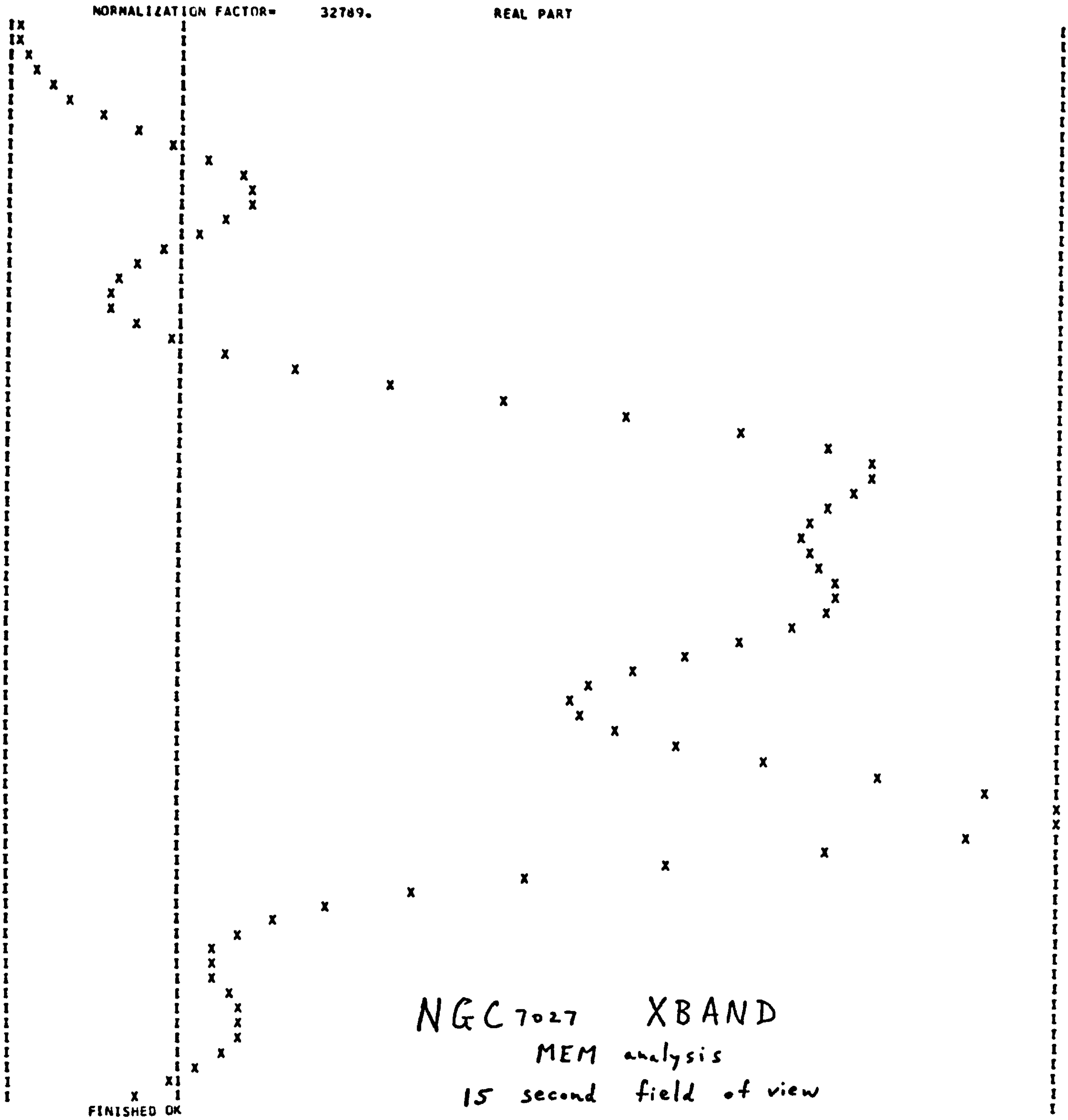


Figure 2



NGC 7027 XBAND
MEM analysis
15 second field of view

Figure 3