PART V

MAXIMUM ENTROPY IMAGE RECONSTRUCTION

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

RESOLUTION ENHANCEMENT; THE "MAXIMUM ENTROPY METHOD" AND THE "HIGH RESOLUTION METHOD"

(Invited paper)

C. van Schooneveld
Leiden Observatory,
Leiden, the Netherlands
and
Physics Laboratory NDRO,
The Hague, the Netherlands.

1. INTRODUCTION

1.1 Resolution enhancement and a-priori knowledge.

In image formation from coherence functions we start with a set of observed coherences (visibilities) $R'(u_k, v_k)$, measured at locations u_k, v_k in the u,v-plane:

$$R'(u_k, v_k) = R(u_k, v_k) + r(u_k, v_k) \; ; \quad k = -M, .., M. \tag{1}$$

R(u,v) is the object's coherence function (visibility function, covariance function) and $r(u_k, v_k)$ are random measurement errors. In conventional image formation we apply a weighted fourier transformation to eq 1 and we get

$$B'(l,m) = \{G(l,m) \circledast B(l,m)\} + b(l,m), \tag{2}$$

where B(l,m) is the brightness distribution of the object and B'(l,m) is the image. G(l,m) is a beampattern depending on the locations u_k, v_k and on the weights applied to them. The component b(l,m) is the random error in the image, corresponding to $r(u_k, v_k)$.

The locations u_k, v_k are usually limited to some central area A of the u,v-plane. Consequently the high-order fourier components of B(l,m), which correspond to coherences outside A, are missing from the image B'(l,m). The result is a loss of resolution.

The effect of resolution loss can be controlled - to some extent - by using suitably tapered weights, but this method does not recover any of the object's missing fourier components. A better job can be done if it is possible to operate on B'(l,m) (or on $R'(u_k, v_k)$) in such a way that part or all of these components are retrieved.

At first glance the problem of inverting eq. 2, i.e. finding B( )

C. van Schooneveld (ed.), Image Formation from Coherence Functions in Astronomy, 197- 218.

from B'( ), looks like an ordinary deconvolution problem, but there is a difference. In a standard deconvolution problem the object's high frequency components are attenuated by a finite factor and are subsequently corrupted by the measurement errors. A solution is then sought within the class of linear operations on B'(1,m) (Wiener filters or modified inverse filters). In eq. 2, however, the high-order fourier components of the object were never observed. They have therefore effectively been attenuated down to zero. As a consequence, any inversion scheme that performs the required task must be non-linear or at least linear space-variant. No linear, space-invariant operation on B'(1,m) can create fourier components that are not already present!

The key to the inversion of eq. 2 is to use a-priori assumptions concerning the object B(1,m) or, better, a-priori knowledge. The assumptions must provide a deterministic coupling between the missing high-order fourier components and the available low-order ones. Eq. 2 can then be solved, for example, by fitting an a-priori constrained object model to the data B'(1,m). Equivalently, in the language of eq. 1, our assumptions must establish a deterministic coupling between the observed coherences inside the area A and the unobserved ones, outside A, thereby allowing an extrapolation of the coherence function. Two questions are important for the success of such an attempt at resolution improvement:

1. Are the assumptions valid for the object under consideration?
2. Is the coupling mechanism sufficiently powerful to withstand the random measurement errors?

Some frequently used assumptions concerning B(1,m) are:

1. Non-negativeness : $B(1,m) \geq 0$
2. Limited extent : $B(1,m) = 0$ outside some closed contour in the 1,m-plane.
3. A variety of assumptions about the structure of the object, all of them characterized by a parametrization of B(1,m) (i.e. the continuous function B(1,m) is known if the values of N parameters have been found).

Based on such assumptions a number of resolution enhancement methods has been proposed (Frieden 1975). We shall discuss two of the better developed ones and we shall try to analyze their underlying principle. They are Burg's Maximum Entropy Method (MEM) and Capon's High Resolution Method (HRM), also erroneously known as Maximum Likelihood Method (MLM). In ch. 2 we repeat an argument from temporal spectrum analysis to show that the MEM, sometimes said to be "maximally non-committal" with regard to the observations, relies in fact on a structural assumption about the object, an assumption which may or may not be valid. The HRM, discussed in chapter 3, was originally developed for adaptive detection of point sources against an unknown and non-uniform radiation background, but Capon realized that the method yields improved resolution when used as an estimator of brightness distributions. We show that the HRM can be interpreted as a positive-constrained reconstruction method.

At this point we comment briefly on methods which rely on the assumption of limited object extent. Historically they were the first enhancement methods proposed and they are known as "superresolution" or "superdirectivity". (The temporal counterpart (section 1.3) is known as "superselectivity".) The idea is that an object of limited size produces an analytic coherence function. This implies that a set of coherence observations, very densely spaced, but extending over only a small area A in the u,v-plane, can be extrapolated to an area many times larger. In this way we can obtain beampatterns G(l,m) with a peakwidth inversely proportional to the number of u,v-samples, rather than inversely proportional to the dimensions of the area A. Physically this is achieved by making the wavefront from the desired steering direction (l_o, m_o) cancel to a high degree, while obtaining an even higher degree of cancellation for wavefronts from the other directions. However, this cancellation mechanism is also the source of the method's weakness: an extreme sensitivity to internal noise like thermal noise in the array sensors, quantization and round-off noise in digital processors, etc. By sensitivity we mean that a superresolving system, designed under the assumption of no internal noise, is completely swamped by such noise if it is indeed present. If, conversely, the internal noise is taken into account in the design, the system automatically reverts to a near-conventional processor and almost no superresolution occurs. In fact the assumption of finite object size is too weak to cope with the measurement errors, when unsupported by other assumptions. Superresolution can be ruled out as a practical tool for high-performance imaging systems. Conclusions of this sort were reached in the fields of E.M. antennas (Gilbert and Morgan 1955, Taylor 1955), accoustics (Cox 1973, Schultheiss 1977), radio astronomy (Cole 1973) and optics (Frieden 1975). We note in passing that other enhancement methods, for example the HRM, have the same flaw in their character - cancellation type processing - and special precautions must sometimes be taken to suppress the danger involved (Cox 1973, Owsley 1973, Griffiths and Hudson, 1977).

1.2 Simplification to 1 dimension.

We shall base our discussions on the 1-dim. equidistant case, i.e. line-shaped arrays with constant baseline increment Δ. The advantage is that we obtain simple closed-form solutions which reveal the mechanism of the method. Apart from this, 1-dim. spatial imaging is of practical importance, for instance in underwater acoustics where long horizontal arrays are used to map the radiation intensity as a function of azimuth. One-dim. strip scans in radio astronomy are another example. Standard radio astronomy images, however, are 2-dimensional. Extension of our discussion to 2-dim. arrays is possible, as well as to non-uniform sampling. Qualitatively the properties remain the same but a good deal of the simplicity of the equations is lost.

We use a fourier notation adapted to a (1-dim.) object B(1) of finite extent: $B(1) = 0$; $|1| \geq W$. (This is merely a matter of notation; we do not exploit the limited size of the object in the sense of superresolution.) B(1) can then be written as a fourier series on an interval

of length $1/\Delta = 2W$:

$$B(l) = \Delta . \sum_{k=-\infty}^{\infty} R_k . \exp - 2\pi jkl\Delta \quad ; \quad -W < l < W \quad ; \quad \Delta = 1/2W, \qquad (3)$$

where the R_k are samples of the coherence function R(u) at equal spacings Δ:

$$R_k = R(k\Delta) = \int_{-W}^{W} B(l) . \exp 2\pi jlk\Delta . dl \quad ; \quad k = -\infty, .., \infty . \qquad (4)$$

With an array of finite length the observed coherences are limited to k = -M, .., M. In conventional image reconstruction eq. 3 is then replaced by a truncated (and weighted) fourier transform:

$$B'(l) = \Delta . \sum_{k=-M}^{M} (R_k + r_k) . g_k . \exp - 2\pi jkl\Delta , \qquad (5a)$$

whereas image enhancement methods try to get hold of the missing coherences, $|k| > M$, via an a-priori coupling mechanism. The r_k in eq. 5 are the measurement errors and the g_k are the weights. Eq. 5a can finally be written in the form of eq. 2:

$$B'(l) = \int_{-W}^{W} G(n-l) . B(n) . dn + b(l), \qquad (5b)$$

where G( ) is the fourier transform of the weights $\Delta.g_k$.

1.3 The space-time analogy.

Resolution enhancement applies equally well to temporal spectrum analysis of random signals as to (pseudo - monochromatic) spatial imaging with a line array or a plane array. Some spatial enhancement methods were first conceived with an application to temporal spectrum analysis in mind, and conversely. Both techniques are analogous ; we briefly recall the analogy below (table 1, fig. 1).

In temporal spectrum analysis we have a random signal s(t) which is sampled at intervals Δ seconds. We try to measure the power density spectrum B(f): the distribution of the average signal power over the temporal frequencies f. In spatial imaging we start with the random aperture illumination $\Delta(x,t)$, the (complex) amplitude of the E.M. field along the aperture axis x at one time instant t. We measure B(l), the distribution of the average power of illumination over the spatial frequencies l. Spatial frequency l corresponds to angular source direction ϕ via $l = \sin \phi / \lambda$. Actually, l is the projected wavenumber of a wavefront from direction ϕ.

In spectrum analysis we store a block of M + 1 samples s_o, .., s_M in a shift register and we calculate the cross-products $s_i\, s_{i+k}$. Averaging over many datablocks produces covariance estimates R'_{-M} , .., R'_M which are processed to give a spectral estimate $\hat{B}(f)$. In the spatial

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

TIME	SPACE
. random signal: s(t)	random aperture illumination: s(x) ("frozen" at one time instant)
. time series after sampling at intervals of Δ sec.: .., s_k, ..	space series after sampling at distances of Δ meter: .., s_k, ..
. covariance function: $R_k = E\{s_i s_{i+k}\}$	coherence function: $R_k = E\{s_i^* s_{i+k}\}$
. power density spectrum: B(f)	brightness distribution: B(l)
f: temporal frequency (sec^{-1})	l: spatial frequency ($meter^{-1}$) (projected wavenumber)
. fourier-relation $B(f) \leftrightarrow R_k$:	fourier relation $B(l) \leftrightarrow R_k$:
"Wiener-Khintchine"	"v. Cittert-Zernike"

Table 1, space-time analogy.

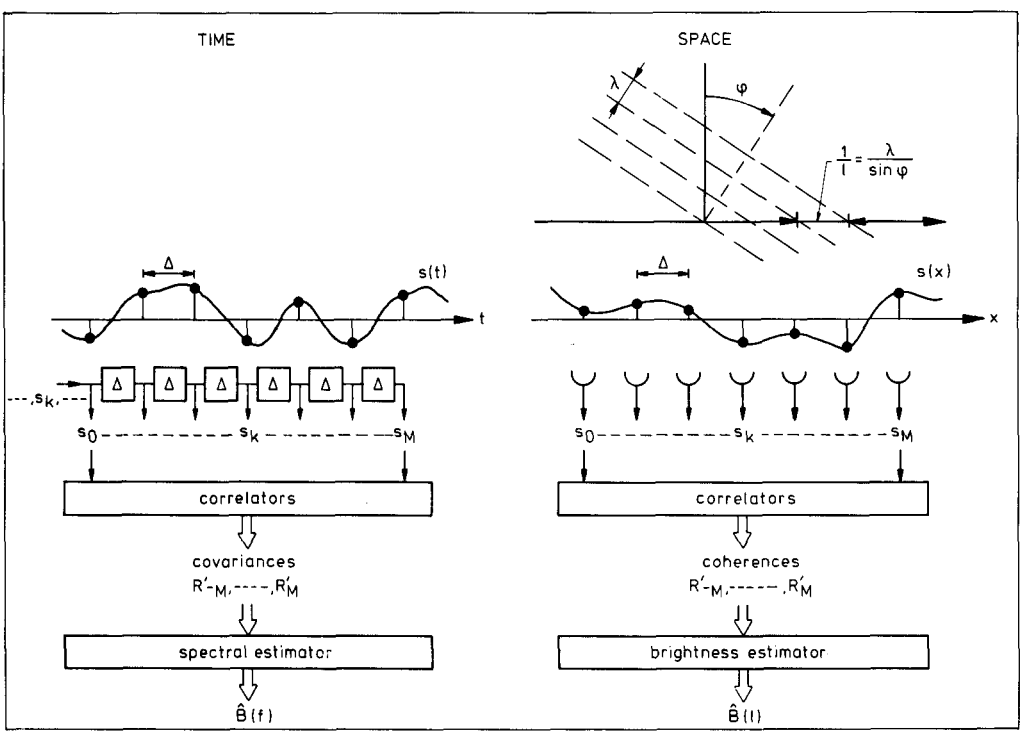


Fig. 1. Temporal spectrum analysis (left) and spatial imaging (right).

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

case the block of M+1 illumination samples, obtained at one time instant t_1, is multiplied to yield crossproducts $s_i^* s_{i+k}$. They are averaged over many successive times t_1, t_2, t_3, .. and the coherence estimates $\hat{R}_{-M}$, .., $\hat{R}_M$ are transformed to an estimated brightness distribution $\hat{B}(l)$.

There is one minor difference: power density spectra are symmetric functions of frequency; spatial objects are not. In order to distinguish positive and negative source directions in the spatial case, we must use a complex-valued illumination s(x). The result is a complex hermitic coherence function. In the temporal case the covariance function is real-valued and symmetric.

2. MAXIMUM ENTROPY METHOD (MEM)

The main purpose of this chapter is to repeat an argument from spectrum analysis (Burg 1967; van den Bos 1971) to show that MEM fits an a-priori constrained object model to the observed coherences.

2.1 MEM and Markov (or autoregressive) processes.

Following Burg (1967) we start in this section with an application of the MEM to spectrum analysis of a random time series .., s_k, .. . We have M+1 observed covariances R_k, k = 0, .., M (or k= -M, .., M since $R_k = R_{-k}$), spaced at Nyquist distances $\Delta = 1/2W$ (eq. 4). We neglect, until further notice, the measurement errors r_k. Burg selected as a spectral estimate the function $\hat{B}(f)$ which maximizes the entropy of the process .., s_k, .. under the condition of consistency with the observations, i.e.

$$\int_{-W}^{W} \hat{B}(f) \cdot \exp 2\pi jfk\Delta \cdot df = R_k \; ; \quad k = -M, .., M \qquad (6)$$

The entropy of the process is a measure for the randomness, or the unpredictability, of .., s_k, .. . Maximum entropy means that a value s_k is maximally unpredictable from its entire past, s_{k-1}, s_{k-2}, .. , $s_{k-\infty}$, given the (M+1) observed covariances. This implies that the first M previous samples, s_{k-1}, .., s_{k-M}, are all that is needed to predict s_k and that the availability of earlier samples s_{k-M-1}, ..., $s_{k-\infty}$ does not improve predictability, despite the fact that they are correlated with s_k. Given the observed R_k's there is only one power density spectrum which belongs to such a "least predictable process"; this particular spectrum is taken as the MEM estimate $\hat{B}(f)$. Observe that the MEM criterion is a subjective one; it is equivalent to the a-priori-assumption of a "least predictable signal" .., s_k, .. . There is no evidence for this in the observed covariances themselves!

Solutions can be obtained in many different ways (Survey papers: Ulrych and Bishop 1975, Makhoul 1975). One approach is to express the entropy of the process in its spectrum (Shannon and Weaver 1949):

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

$$H \doteq \frac{1}{2W} \int_{-W}^{W} \ln B(f) \, . \, df, \tag{7}$$

and to maximize (7) subject to (6). A second approach (van den Bos 1971) is based on the theory of (wide sense) Markov processes. (Papoulis 1965). Our "least predictable signal", consistent with M+1 covariances, is a Markov process. Such processes can only be generated by an Mth order autoregressive filter (A.R. filter), with unit variance uncorrelated noise samples .., w_k, .. at the input (fig. 2). The output covariance function, although infinitely long, is determined by M+1 parameters, b_o, a_1, .., a_M. The relevant equations are:

$$\begin{pmatrix} R_o & \cdots & R_M \\ \vdots & & \vdots \\ R_M & \cdots & R_o \end{pmatrix} \begin{pmatrix} 1 \\ -a_1 \\ \vdots \\ -a_M \end{pmatrix} = \begin{pmatrix} b_o^2 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \text{and} \tag{8a}$$

$$R_k = \sum_{i=1}^{M} a_i \, R_{k-i} \; ; \quad k = M+1, \, M+2, \, \ldots, \, \infty \; . \tag{8b}$$

Given R_o, .., R_M we can solve eqs. 8a for b_o^2, a_1, .., a_M. This implies that we fit the Mth order Markov, or A.R., model to the data. We can then extrapolate the covariance function via eq. 8b and take the fourier transform to obtain the estimate $\hat{B}(f)$. Observe that eq 8b represents the coupling mechanism discussed in sec. 1.1. An equivalent method, less elaborate, is to solve for b_o^2, a_1, .., a_M and to take the AR filter's output spectrum as the estimate:

$$\hat{B}(f) = \frac{b_o^2}{\left| 1 - \sum_{i=1}^{M} a_i \, . \, \exp - 2\pi j f i \Delta \right|^2} \, . \Delta \tag{9}$$

Eq. 9 has a (2M+1)-term fourier series in the denominator. Thus, $\hat{B}(f)$ contains fourier components up to infinity: many more than the conventional estimate of eq. 5.

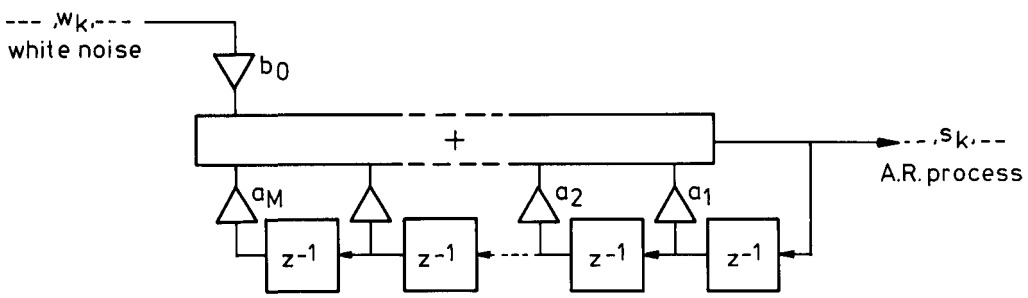


Fig. 2. Mth order AR filter. z^{-1} is the unit delay operator. This is the model, fitted to the data in a MEM spectrum analysis. The implicit assumption is that the spectrum consists only of resonances (poles).

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

The assumption of a "least-predictable" or A.R. or Markov process implies an a-priori assumption concerning the spectrum. Eq. 9 can be factored to give

$$\hat{B}(f) = \frac{b_o^2}{\left|(\exp 2\pi jf\Delta - p_1) \,..\, (\exp 2\pi jf\Delta - p_M)\right|^2} .\Delta \quad , \qquad (10)$$

where $p_1, \ldots, p_M$ are the (real valued or conjugate paired) roots of $(z^M - a_1 z^{M-1} - \ldots - a_{M-1} z - a_M)$. The roots are the poles, or resonances, of $\hat{B}(f)$: a resonance occurs if a frequency f brings one of the terms exp $2\pi jf\Delta$ close to a pole p_i. Thus, the MEM estimate consists of resonances only ("all-pole-spectrum"). A general spectrum contains absorptions (zeroes) as well as resonances (poles). It would be obtained at the output of a filter with feed-forward branches in addition to feedback branches (fig. 3).

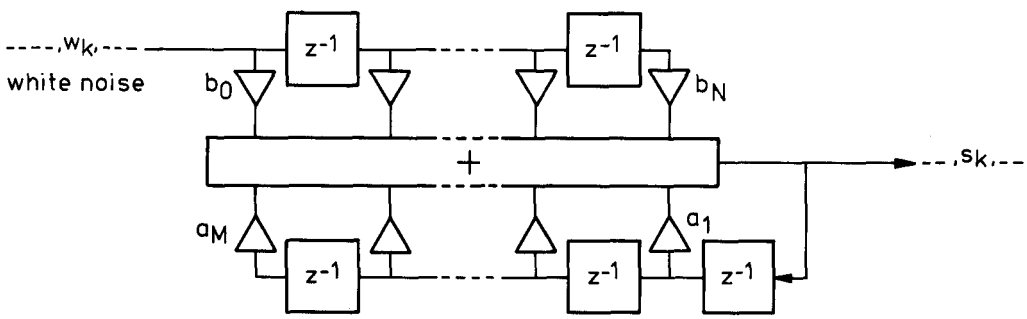


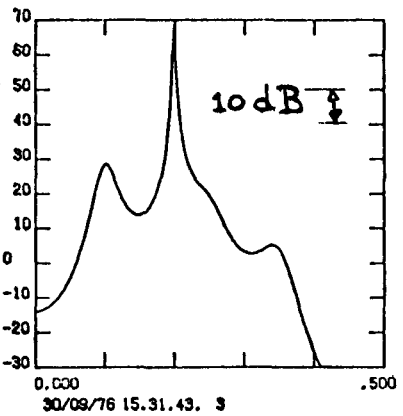
Fig. 3. General digital filter. This model should be used in the case of resonances (poles) and absorptions (zeroes).

For a mixed pole-zero spectrum the model of fig. 3 should be fitted to the data, rather than an A.R. model, and the solution would no longer correspond to an entropy maximization. Application, in this case, of the MEM leads at most to an approximation and it should be investigated whether this is a better one than the conventional fourier transform of eq. 5, or not. In the case of an all-zero spectrum the situation becomes surprisingly simple: no model fit is required since the covariance function R_k assumes a finite length. Under the proviso that all non-zero covariances have been observed there is no need for resolution improvement; the conventional fourier transform (eq. 5, unweighted) is optimum! Succesful applications of MEM to the detection and resolution of sharp spectral lines were reported in the past (references in Ulrych and Bishop 1975 and in Lacoss 1971). Spectral lines in temporal frequency analysis - and point sources in spatial imaging - are limit cases of resonances and are therefore marginally valid objects for the MEM.

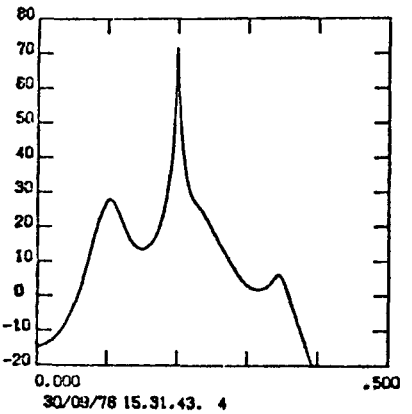
https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

2.2 Example

Fig. 4a shows an object spectrum B(f) of 20 poles (no zeroes). It was deliberately designed to have a narrow peak, some weak structure around the peak, and a dynamic range of > 100 dB. A piece of signal .., s_k, .. of 1024 points was used to determine the first 21 covariances, R_o, .., R_{20}. Fig. 4c shows the conventional estimate $\hat{B}(f)$ (eq. 5; g_k = gaussian taper): all spectral details are lost. The MEM estimate $\hat{B}(f)$, based on the same 21 covariances, is shown in fig. 4b. The deviations from the nominal spectrum B(f) are less than 2dB; they are due to the errors r_k caused by the finite (1024) signal duration (section 2.3).

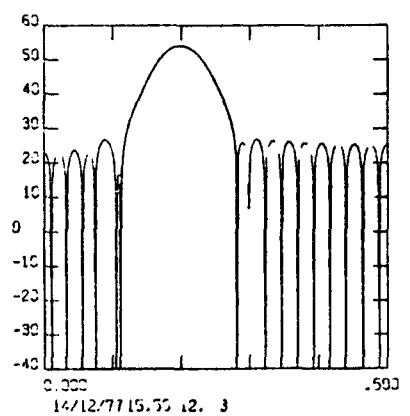


a) Spectrum B(f).



b) MEM estimate $\hat{B}(f)$(eqs 8,9).

Fig. 4. Maximum Entropy Method (MEM). The spectrum B(f) belongs to a 20th order Markov (or A.R.) process. The estimates $\hat{B}(f)$ are based on 21 observed covariances. The conventional estimate suffers an important resolution loss. The lost resolution is recovered by the MEM estimate.



c) Conventional estimate $\hat{B}(f)$ (eq. 5).

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

Note that the conventional estimate (fig. 4c) could have been plotted on a wide f-grid without serious loss of detail. The MEM estimate, on the other hand, requires a dense plotting grid if the benefits of resolution enhancement are to be retained.

2.3 Measurement errors.

Actual observations of R_k are corrupted by errors r_k which propagate into the MEM estimate. The problem is particularly troublesome when B(f) has a large dynamic range. A general theory of error propagation is still failing but MEM does not seem to be unduly sensitive to random errors, provided that the problem of order estimation can be solved satisfactorily.

Up to now we have tacitly assumed that the order of the AR model is equal to the number of observed covariances, which implies that M+1 degrees of freedom are fitted to M+1 observations. Clearly we must not use too few degrees of freedom (resolution loss) but we must also avoid the danger of using too many (sensitivity to random errors) and we are therefore in need of algorithms for estimating the order of the model to be fitted. Various such methods are being developed, both for AR order estimation (Lacoss 1977, Jones 1976, Ulrych and Bishop 1975, Parzen 1974) and for model fitting in general.

Akaike's "Information Criterion" method (Akaike 1974) is an example from the 2nd class. An application is described by Ishiguro and Ishiguro (this volume). The method is a generalization of "naive" order estimation where one studies the behaviour of S, the sum of the squared residuals of a fitting procedure, as a function of N, the number of degrees of freedom. For small N the value S(N) drops rapidly as N increases. For large N the value of S decreases slowly and linearly until S = 0 when N becomes equal to the number of observations, M+1. The estimated order is the value of N where the linear behaviour sets in. The use of a value $N_{opt} < M+1$ leads, of course, to non-zero residuals. Consequently, the consistency conditions of eq. 6 must be relaxed.

Various authors have relaxed the consistency equations by maximizing eq. 7 under the condition

$$\frac{1}{M+1} \sum_{k=0}^{M} \left| R'_k - \int_{-W}^{W} \hat{B}(f) \cdot \exp - 2\pi j f k \Delta \cdot df \right|^2 = \sigma^2, \qquad (11)$$

where σ^2 is the variance of the r_k errors and R'_k are the observations (eq. 1). (Ables 1974, Wernecke and d'Addario 1977, Wernecke 1977, Gull and Daniell 1978). The advantage is that many constraints are replaced by a single one. In addition one obtains an "error-tolerant" estimate $\hat{B}(f)$ with an entropy value larger than obtained under the conditions of eqs. 6. Newman (1977) shows that the structure of the solution $\hat{B}(f)$ remains indentical to eq. 9, which suggests that the number of degrees of freedom is still equal to the number of observations. It would be

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

interesting to compare the results of this approach with the results of order estimation methods.

2.4 One-dimensional spatial imaging.

Our discussion remains valid for 1-dim. spatial imaging with constant baseline increments Δ along a linear axis x, the aperture axis, provided that we use complex, hermitic coherences R_k. The underlying MEM assumption is, that the spatial series .., s_k, .. (E.M. field samples along the array axis) has a spatial AR- or Markov-structure, i.e. that M spatial samples are sufficient to predict the next spatial sample and that other samples, if available, would be useless for spatial prediction.

2.5 Two-dimensional spatial imaging.

The concept of maximizing eq. 7 under the constraints of eq. 6, or eq. 11, can be extended to 2-dimensional sets of coherence observations in the u,v-plane. Applications to radio astronomy were described by Ponsonby (1973), Wernecke and d'Addario (1977) and Gull and Daniell (1978). For simplicity we restrict ourselves to observations on a rectangular grid with spacings Δ in both dimensions:

$$R'_{ki} = R(k\Delta, i\Delta) + r_{ki} \; ; \quad k,i = -M, \ldots, M \; ; \quad R_{ki} = R^{*}_{-k,-i} \tag{12}$$

Newman (1977) showed that the estimated brightness assumes the form

$$\hat{B}(l,m) = \frac{b_o^2}{\left| 1 - \sum_{\substack{g,h=0 \\ g+h\neq 0}}^{M} \sum a_{gh} \cdot \exp - 2\pi j(gl + hm)\,\Delta \right|^2} \cdot \Delta^2 \; , \tag{13}$$

i.e. the same structure as eq. 9. He derived eq. 13 in two ways: 1) maximization of the entropy $\int \ln \hat{B}(l,m)\, dl\, dm$ under consistency constraints, 2) via a minimum prediction error method. He concludes from the similarity of the results that the link between an AR model fit and the MEM extends to problems of 2 dimensions.

The implication is that the underlying a-priori assumption - a least predictable random process - remains valid in two dimensions. However, our process is a 2-dim. space series .., s_{gh}, .., the E.M. field amplitude at the aperture sample points and we must explain what is meant by a "least predictable process, subject to a set of observed coherences". This takes us into the area of 2-dim. Markov processes (Woods 1972, Jain 1977).

A process .., s_{gh}, .. is Markov if a sample value s_{g_o,h_o} can be predicted in terms of a limited number of samples in the

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

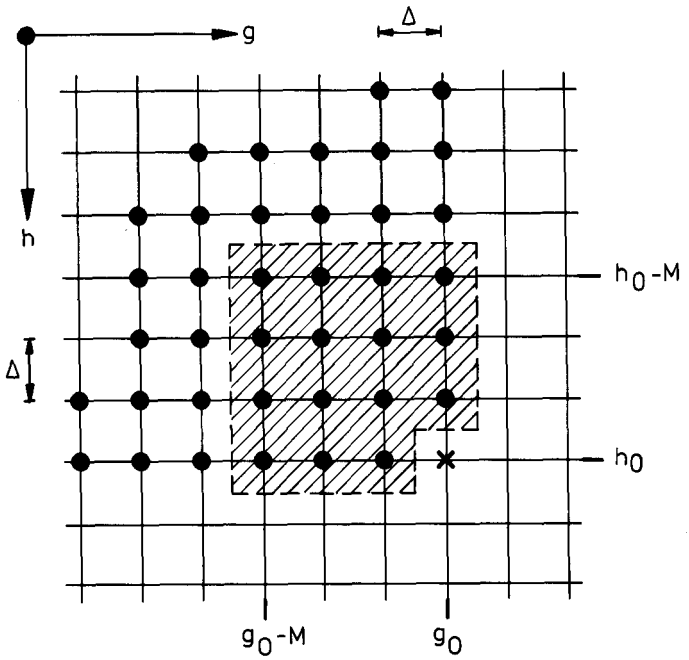


Fig. 5. Causal 2-dim. Markov process; order MxM. Prediction at g_o, h_o, based on all points $g \leq g_o$, $h \leq h_o$ $(g + h \neq 0)$ gives the same result as a prediction, based on the points $g_o - M \leq g \leq g_o$, $h_o - M \leq h \leq h_o$ $(g+h \neq 0)$. This is the implicit model used in a 2-dim. MEM.

vicinity of the point g_o, h_o. The other points are superfluous for prediction. Just as in the 1-dim. case, the MEM seems to be exclusively related to causal Markov processes. Fig. 5 illustrates a 2-dim. causal Markov process (Jain 1977). The vicinity, sufficient for prediction of s_{g_o,h_o} is "to the left and above" of the point g_o, h_o (hatched area); all other points in the upper left quadrant $g \leq g_o$, $h \leq h_o$ are superfluous. Causal Markov processes obey a difference equation analogous to the diagram of fig. 2, with coefficients a_{gh} $(g, h = 0, .., M; g+h \neq 0)$ and b_o. The coefficients can be solved from the observed coherences of eq. 12 via relations analogous to eqs. 8. As in the 1-dim. case, the brightness distribution is finally obtained as the power transfer function of the difference equation, which leads to eq. 13.

Thus, the conclusion seems justified that a 2-dim. MEM relies on fitting a 2-dim. Markov process to the observations.

2.6 Another entropy expression.

The preceding sections were based on the entropy expression of eq. 7: $H_1 \doteq \int\int \ln \hat{B}(l,m).dl\ dm$. We note that H_1 is a property of the random process .., s_k, .., related to its (un)predictability. It is not

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

a measure for the likelihood of the estimate $\hat{B}(1,m)$, unless of course we have the a-priori certainty that .., s_k, .. is an all-pole-process.

More recently a second entropy expression was introduced: $H_2 = \iint \hat{B}(1,m).\ln \hat{B}(1,m)\ .\ dl\ dm$. Maximization of H_2 under consistency conditions is said to give a measure for the likelihood of $\hat{B}(1,m)$ (Frieden 1975, Gull and Daniell 1978, Kikuchi and Soffer 1977, various papers in this volume).

When using H_2 instead of H_1 the connection with Markov processes is broken and our preceding discussion does not apply anymore. Still, H_2 gives improved resolution, implying that the coherence function has been extrapolated, in one way or another. It would be interesting to find out if H_2 can also be interpreted as a model fit. In other words: what are the conditions on the object B(1,m) which make a H_2-coherence extrapolation a legitimate one?

3. HIGH RESOLUTION METHOD (HRM)

We analyze the mechanism of Capon's High Resolution Method (HRM) - also known, erroneously, as Maximum Likelihood Method (MLM) - and we indicate two areas which require further investigation.

3.1 Imaging with a beamformer-wattmeter combination.

We start with an array of M+1 sensors and we assume that all outputs s_0, .., s_M (aperture illumination) are simultaneously available for processing. The sensors may be distributed over a linear or a plane aperture, but for simplicity we shall discuss a 1-dim. line-shaped array with equal sensorspacings Δ (fig. 6). Note that the condition of simultaneous accessibility of all sensor outputs is not fulfilled in 2-dim. earth rotation aperture synthesis; we return to this point in sec. 3.6.

Fig. 6 shows the traditional array processor, consisting of a beamformer and a wattmeter. The beamformer adds the instantaneous sensor outputs s_i via weights $a_i^*(l_o)$ so as to steer an antenna beam in the desired direction l_o:

$$x = \sum_{i=0}^{M} a_i^*(l_o)\ .\ s_i\ . \tag{14}$$

The wattmeter estimates the power, incident upon the array from direction l_o, by squaring x(t) and averaging over a finite time T. Note that beams can be steered simultaneously in many different directions l_o by using different sets of weights for each beam. Neglecting the measurement error caused by a finite integration time T, the average output power is given below, where E denotes expectation:

$$\hat{P}(l_o) = E\left\{\left|\ \sum_{i=0}^{M} a_i^*(l_o)\ .\ s_i\ \right|^2\right\}. \tag{15}$$

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

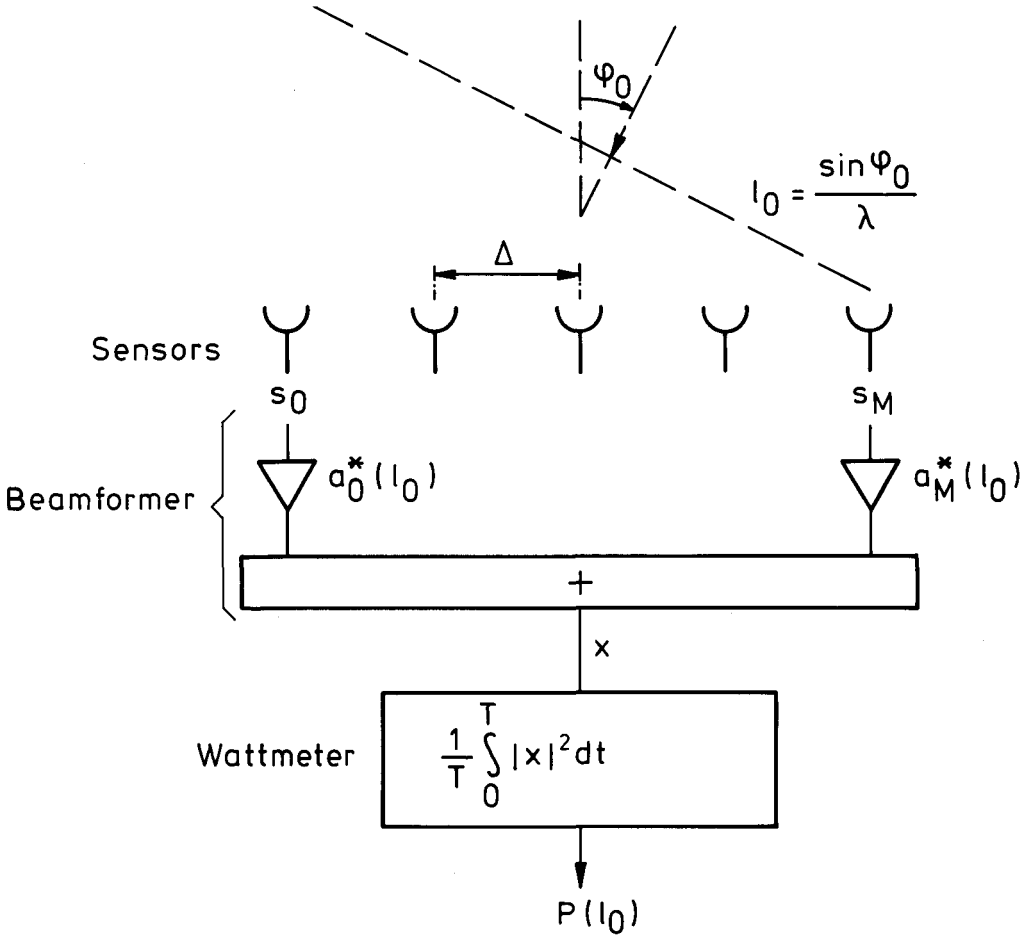


Fig. 6. Beam former + wattmeter; a traditional imaging system when all sensor outputs are simultaneously available for processing. l_o is the beamsteering direction. Using a parallel processor, beams can be steered in many different directions.

Invoking the definition of the coherence function, $E\{s_i^* s_{i+k}\} = R_k$, and the van Cittert-Zernike theorem, $R_k = \int B(l) \cdot \exp 2\pi j k l \Delta \cdot dl$, this can be written in various other forms:

$$\hat{P}(l_o) = \sum_{i,m=0}^{M} \sum a_i^*(l_o) \cdot a_m(l_o) \cdot R_{i-m} \tag{16}$$

$$= \sum_{k=-M}^{M} w_k^*(l_o) \cdot R_k \tag{17}$$

$$= \int W(l, l_o) \cdot B(l) \cdot dl, \tag{18}$$

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

where

$$w_k^*(1_o) = \sum_{i=0}^{M} a_i^*(1_o) \cdot a_{i-k}(1_o) \tag{19}$$

$$W(1, 1_o) = \sum_{k=-M}^{M} w_k^*(1_o) \cdot \exp 2\pi j1k\Delta = \left| \sum_{i=0}^{M} a_i^*(1_o) \cdot \exp 2\pi jli\Delta \right|^2 \tag{20}$$

All integrals above and below are over the source interval $[-W = -1/2\Delta$ $W = 1/2\Delta]$(eqs. 3,4). Eq. 18 expresses the output power in the familiar form of the brightness distribution B(1), convolved with $W(1, 1_o)$, the antenna beampattern (BP).

Observe that eq. 18 allows the use of differently shaped BP's for different steering directions, by selecting specific sets of weights $a_i^*(1_o)$ (or $w_k^*(1_o)$) for each 1_o. This opens the possibility to adapt the imaging system to the brightness distribution. Although $W(1, 1_o)$ is limited to 2M+1 fourier components (eq. 20), we can realize quite arbitrary, and even asymmetrical, BP's. We shall take advantage of this feature by minimizing the power, picked up by our antenna beam, from directions other than 1_o. Or, in the jargon of spectrum analysis, we shall minimize the "leakage" from object coordinates 1 to the steering direction 1_o.

Eq. 18 holds also the potential of resolution improvement. Suppose that we divide eq. 18 by the effective width of the beampattern, say L, and that we regard $\hat{P}(1_o)/L$ as a brightness estimate $\hat{B}(1_o)$. In that case eq. 18 relates the estimate to the object via a shift-variant operation, contrary to a conventional image reconstruction which produces a shift-invariant operation, similar to eq. 5b:

$$\hat{B}(1_o) = \int G(1-1_o) \cdot B(1) \cdot d1. \tag{21}$$

(The conventional result can be obtained by putting $w_k^*(1_o) = g_k \cdot \exp - 2\pi jk1_o\Delta$ in eq. 17; compare with eq. 5a; the pattern G ( ) in eq. 21 is the fourier transform of g_k, k = -M, .., M.) In eq. 21 the number of fourier components in $\hat{B}(1)$ is limited by the number in G(1), i.e. k = -M, .., M. In eq. 18, however, the range of fourier components in $\hat{B}(1)$ may exceed this number, i.e. it may exceed the limit set by the length of the antenna array, and we may hope for a resolution improvement.

3.2 Capon's High Resolution Method (HRM).

Capon (1969; Capon and Goodman 1970) imposed the following conditions on the beampattern:

$$1.\ W(1, 1_o) \geq 0; \quad \text{all } 1, \text{ any } 1_o \tag{22}$$

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

$$2.\ W(l_o, l_o) = 1 \tag{23}$$

$$3.\ \hat{P}(l_o) = \text{mimimum}, \tag{24}$$

and he proposed to regard the minimized output power $\hat{P}(l_o)$ as an estimate of the power radiated by the object B(l) in a narrow region around l_o. What is the significance of these conditions?

1. By requiring a non-negative beampattern we have the certainty of a positive estimate $\hat{P}(l_o)$. Hence the method relies on the a-priori assumption of a positive brightness distribution (sec. 1.1). (For a processor of the type of fig. 6 the positivity assumption is already incorporated because of the presence of the wattmeter.)

2. Eq. 23 normalizes the beampattern's power sensitivity in the steering direction to unity. Radiated power from directions close to l_o is measured correctly.

3. By minimizing the average output power $\hat{P}(l_o)$ under the constraints 22 and 23 we hope to minimize the power picked up via sidelobes of the BP (minimization of "leakage"; sec. 3.1). We shall see that this leads to a BP where the sidelobes have arranged themselves "inversely" with respect to the brightness distribution: low sidelobes where B(l) is large, and conversely.

Eqs. 22, 23, 24 are solved very easily for a beamformer system. No special precautions are required to satisfy eq. 22; the BP is always positive because of eq. 20. Further, eqs. 20 and 16 are used to express $W(l_o, l_o)$ and $\hat{P}(l_o)$ in terms of the weights $a_i^*(l_o)$. The minimization is then performed via a Lagrange multiplier and we obtain an optimum set of weights $a_i^*(l_o)$ which can be resubstituted in eq. 16. We shall skip the calculations (Lacoss 1971, Cox 1973 (2), Capon 1969). The result is:

$$\hat{P}(l_o) = \frac{1}{\sum_{i,k=0}^{M} Q_{ik} \cdot \exp - 2\pi j(i-k)\, l_o\, \Delta}, \tag{25}$$

where Q_{ik} are the elements of the inverse, Q, of the coherence matrix R, which consists of the elements $R_{ik} = E\{s_i^* s_{i+k}\}$:

$$R = \begin{pmatrix} R_{oo} & R_{o1} & \cdots & R_{oM} \\ R_{10} & R_{11} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ R_{Mo} & \cdots & \cdots & R_{MM} \end{pmatrix} ; \quad Q = R^{-1} = \begin{pmatrix} Q_{oo} & \cdots & \cdots & Q_{oM} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ Q_{Mo} & \cdots & \cdots & Q_{MM} \end{pmatrix} \tag{26}$$

(In the present case of an equidistant linear array (fig. 6) we have $R_{ik} = R_{k-i}$). We make the following observations.

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

1. Although we started with a beamformer, requiring the simultaneous availability of all sensor outputs, we end up with an expression (25) that enables us to use observed coherences R_{ik}. In the equidistant case we only need the values R_{-M}, .., R_M since $R_{ik} = R_{k-i}$. Thus, we can start by measuring coherences in the usual way and then substitute them in eqs. 25, 26 to obtain $\hat{P}(1_o)$. Note that this implies the use of slightly incorrect values since the finite correlation time creates an error r_k (eq. 1).

2. Eq. 25 can be written as a fourier series of infinite length (finite length fourier series in the denominator). Hence $\hat{P}(1_o)$ contains fourier components missing from a conventional image (eq. 5a). Equivalently, the coherence function corresponding to eq. 25 has an infinite length and contains many of the unobserved coherences.

3. Unlike the MEM, the HRM solution (25) is not consistent with the original observations: the first M+1 coherences, calculated from eq. 25, differ from the observed R_o, .., R_M. This is simply checked by doing the calculations for a short series.

3.3 Example.

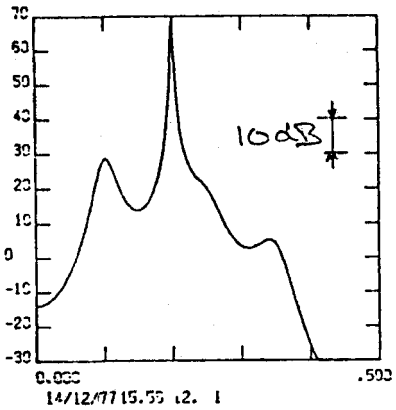
Fig. 7 shows an example of the HRM in temporal spectrum analysis, applied to the same 21 covariances as used in the MEM example of fig. 4. The nominal spectrum B(f) and the HRM estimate $\hat{P}(f_o)$ are depicted in figs. 7a and 7b, respectively. The resemblance is reasonable. Comparison with the conventional estimate of fig. 4c shows a significant enhancement of resolution. Note the discrepancy between the vertical scales of fig. 7a and 7b, caused by the fact that eq. 25 represents power rather than power density.

The plots of fig. 7c present beampatterns $W(f, f_o)$ for 4 different tuning frequencies ("steering directions") f_o, indicated by vertical lines. Observe that $W(f_o, f_o) = 1$, in accordance with eq. 23. Comparison of the BP's with the object spectrum B(f) illustrates the "inverse" behaviour of the sidelobes, in response to the required leakage minimization (eq. 24). The BP's have adapted themselves to the data. The effect is so strong that the main beam of the antenna pattern is no longer clearly recognizable.

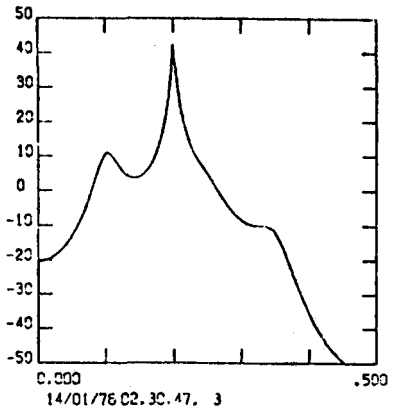
3.4 Discussion.

In sec. 3.2 we derived the HRM estimate $\hat{P}(1_o)$ as the minimum power that can be measured with an all-positive beampattern, normalized to unit sensitivity in the steering direction. Thus, we interprete HRM as a positive constrained reconstruction method which minimizes leakage.

Yet it is amazing that a result of the quality of fig. 7b can be obtained solely on the assumption of a positive object function and one may wonder whether there is perhaps a hidden structural assumption behind the HRM, similar to the MEM assumption. This suspicion is fed by

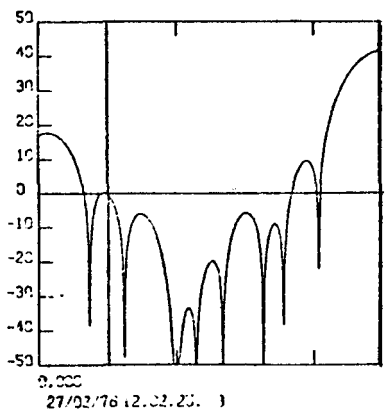


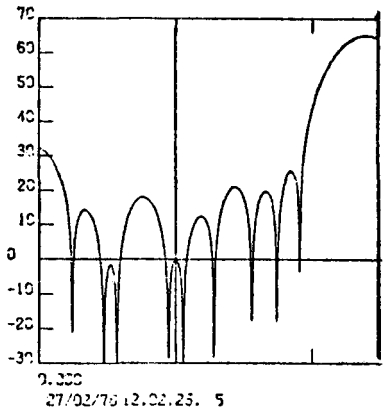
a) Spectrum B(f)

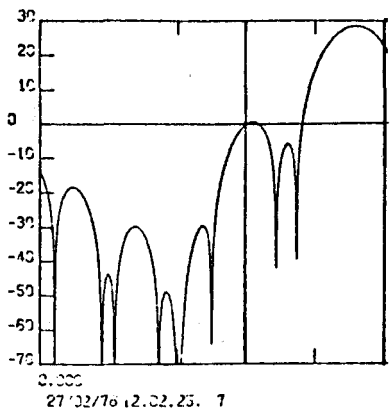


b) HRM estimate $\hat{P}(f)$

c) Beampatterns $W(f, f_o)$ for 4 different steering directions f_o.







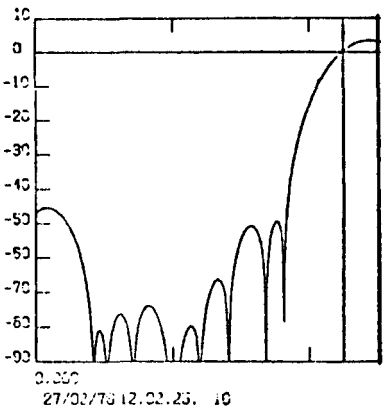


Fig. 7. High Resolution Method (HRM). Same spectrum B(f) as in fig. 4a. Comparison with the conventional estimate (fig. 4c) illustrates the resolution enhancement. Figs. 7c illustrate how the beampatterns adapt in an inverse sense to the object B(f), in order to minimize leakage.

https://doi.org/10.1017/S0252921100074893 Published online by Cambridge University Press

the existence of an algebraic relation between MEM and HRM estimates, as demonstrated by Burg (1972).

3.5 Areas for further study.

There are at least two problems requiring further investigation: 1) intensity calibration and 2) incomplete coherence matrices. The 2nd point is important for application to 2-dim. earth rotation aperture synthesis.

Intensity calibration. We recall that eq. 25 is an estimate of the power radiated by the object within a narrow region around the steering direction; it is not an estimate of the brightness (i.e. density of power) B(l). The HRM has a long history in the area of underwater acoustics where the problem is to detect the presence or absence of a point source $P_s.\delta(l-l_o)$ against an unknown, continuously distributed background radiation B(l). The HRM, in the form of a power estimator, is perfectly suited to this purpose since the constraints 23 and 24 guarantee a maximum signal-to-noise ratio if a point source is indeed present in direction l_o. (These applications go by the name of MLM: Maximum Likelihood Method, a misnomer since the method does not provide an ML estimate of power. Cox 1969, Cox 1973 (1), Cox 1973 (2), Lewis and Schultheiss 1970, Schultheiss 1977.)

Our present purpose is slightly different: we are interested in a density estimate. Hence eq. 25 must be divided by the effective width L of the BP's main beam. Since the BP's may assume strange shapes (fig.7c), and since the shapes differ from one steering direction to the next, the problem deserves some more attention. Various approaches can be taken.

1. An ad-hoc approach is to divide eq. 25 by the effective width L of a conventional unweighted beampattern G(l), to be obtained via eqs. 5.
2. Another option is to use the actual beampatterns $W(l, l_o)$, to be derived from eq. 20, but this poses the problem of defining an effective width for BP's shaped like those of fig. 7c.
3. A 3d solution might be to replace constraint 23 by a modified one:

$$\int W(l, l_o) \cdot C(l-l_o) \cdot dl = 1, \qquad (27)$$

where $C(l-l_o)$ is a positive pulse-type function (e.g. a gaussian), centered at l_o, with an area 1 and an effective width L which could, for instance, be taken equal to a desired resolution width. Observe that eq. 27 reduces to the original constraint 23 if $C(l-l_o)$ is a delta function.

Incomplete coherence matrix. The preceding discussions extend immediately to non-equidistant arrays, distributed over a 2-dim. plane aperture. One can use a beamformer + wattmeter if all sensor outputs are simultaneously accessible or one can start from a previously measured coherence matrix R (eqs. 25, 26).

In the last case, however, the HRM requires that all cross coherences R_{ik} have been measured! This condition is not met in radio astronomy earth rotation aperture synthesis, where the usual observation scheme limits the coherences to those involving the sensor at the origin of the rotating array. This leads at best to the elements in the first row, the first column and the main diagonal of the coherence matrix R. Papadopoulos (1975) applied the HRM to radio astronomical data and he solved the dilemma by setting all unobserved matrix elements equal to zero.

A better solution for the case of an incomplete matrix is to break the link with simultaneously accessible arrays and to rephrase the problem. Suppose that a set of coherences R_{-M}, .., R_M has been observed, distributed arbitrarily in the u,v-plane. The HRM conditions could then read:

$$\begin{aligned} &1.\ \text{Minimize } \hat{P}(1_o) = \sum_{-M}^{M} w_k^*(1_o).R_k \quad \text{under the constraints} \\ &2.\ W(1, 1_o) \geq 0 \\ &3.\ W(1_o, 1_o) = 1\ , \end{aligned} \tag{28}$$

where $W(1, 1_o)$ is the fourier transform of the weights $w_k^*(1_o)$. Unfortunately the solution of this problem is less straightforward than the one of eqs. 22, 23, 24 since a convenient relation of the type 20, automatically satisfying the positivity constraint, is no longer available.

ACKNOWLEDGEMENT

The author is grateful to P.J. Prinsen of the Physics Laboratory NDRO for many dicussions and for the preparation of figs. 4 and 7.

LITERATURE

Ables, J.G.: 1974,"Astron. Astrophys. Suppl." 15, pp. 383.
Akaike, H.: December 1974, IEEE Trans. AC-19, 6, pp. 716.
Burg, J.P.: 1967, Proc. 37th Meeting Soc. Explor. Geophys..
Burg, J.P.: April 1972, "Geophysics", 37, 2, pp. 375.
Bos, A. van den: July 1971, IEEE Trans IT-17, pp. 493.
Capon, J.: August 1969, Proc. IEEE, 57, 8, pp. 1408.
Capon, J., Goodman, N.R.: October 1970, Proc. IEEE, 58, 10, pp. 1785.
Cole, T.W.: 1973, "Astron. and Astrophys." 24, pp. 41.
Cox, H.: 1973, 1, in "Signal Processing", eds. Griffith, Stocklin, van Schooneveld, Academic Press.
Cox, H.: September 1973, 2, "Jour. Acoust. Soc. Am," 54, 3, pp. 771.
Frieden, B.R.: 1975, in "Picture Processing and Digital Filtering", ed. T.S. Huang, Springer.

Gilbert, E.N., Morgan, S.P.: 1955, "Bell System Tech. J." 34, pp. 637.
Griffiths, J.W.R., Hudson, J.E.: 1977, in "Aspects of Signal Processing", pt 1. ed. G. Tacconi, Reidel.
Gull, S.F., Daniell, G.J.: April 1978, "Nature" 272, pp 686.
Jain, A.K.: September 1977, "Jour. Optim. Theory and Applic." 23, 1, pp. 65.
Jones, R.H.: August 1976, "Geophysics" 41, 4, pp. 771.
Kikuchi, R., Soffer, B.H.: December 1977, "J. Opt. Soc. Am." 67, 12, pp. 1656.
Lacoss, R.T.: August 1971, "Geophysics" 36, 4, pp. 661.
Lacoss, R.T.: 1977, in "Aspects of Signal Processing", pt 2, ed. G. Tacconi, Reidel.
Lewis, J.B., Schultheiss, P.M.: 1970, "Jour. Acoust. Soc. Am.", 49, 4 (part 1), pp. 1083.
Makhoul, J.: April 1975, Proc. IEEE, 63, 4.
Newman, W.I.: January 1977, IEEE Trans. IT-23, 1, pp. 89.
Newman, W.I.: 1977, "Astron. Astrophys." 54, pp. 369.
Owsley, N.L.: 1973, in "Signal Processing", op. cit.
Papadopoulos, G.D.: January 1975, IEEE Trans. AP-23, 1, pp. 45.
Papoulis, A.: 1965, Probability, Random Variables and Stochestic Proceses, McGraw-Hill.
Parzen, E.: December 1974, IEEE Trans. AC-19, 6, pp. 723.
Ponsonby, J.E.B.: 1973, Mon. Not. R. Astr. Soc. 164, 0, pp. 1.
Schultheiss, P.M.: 1977, in "Aspects of Signal Processing", pt 1, ed. G. Tacconi, Reidel.
Shannon, C.E., Weaver, W.: 1949, The Mathematical Theory of Communication, Univ. of Illinois Press.
Taylor, T.T.: 1955, IRE Trans. AP-3, 16.
Ulrych, T.J., Bishop, T.N.: February 1975, Rev. Geophys. and Space Phys. 13, 1, pp. 183.
Wernecke, S.J., d'Addario, L.R.: April 1977, IEEE Trans. C-26, 4, pp.351.
Wernecke, S.J.: Sept.-Oct. 1977, Radio Science, 12, 5, pp. 831.
Woods, J.W.: March 1972, IEEE Trans. IT-18, 2, pp. 232.

DISCUSSION

Comment R.H.T. BATES
Do you not think that superresolution can work (using Gerchberg's algorithm, for instance) provided the resolution is only a few Shannon samples? See, for instance, McDonnell and Bates, Ap.J. 208, Sept. 1976, 443.

Reply C. van SCHOONEVELD
The only useful application of superresolution that I know of, has to do with very short antenna's or antenna arrays, i.e. arrays with a length of one or two Nyquist intervals Δ, or less. One can then extrapolate the coherence function over a distance of the order of Δ beyond the physical array limits, before the effect of internal noise becomes prohibitive. Unfortunately, this limits superresolution to antenna

systems where the resolved angle is an appreciable fraction of the field of view. In addition, some useful applications were reported in relation to short end-fire arrays.

Comment J.C. DAINTY
The application of MEM to power spectrum analysis leads to a "log B" maximization implying certain structural constraints on the random process. Astronomers sometimes use "B log B" -MEM- methods. What are the structural constraints in this case?
Reply C. van SCHOONEVELD
I wish I knew! This is an interesting area for further research. The variational problem involved is probably too complicated to lead to a closed-form solution, but it might give a hint about the structure of the solution. This might then lead to a guess about an underlying assumption concerning the process.

Comment L.R. D'ADDARIO
I think that the MLM is not useful for fourier synthesis in radio astronomy. It computes, for each map point, the most likely flux assuming the rest of the map is empty. This is often an unreasonable assumption. Also, the simple matrix inverse does not apply to rotation synthesis data; an instantaneous array with all elements cross-correlated is required.
Reply C. van SCHOONEVELD
I do not agree with your first statement. The MLM, or HRM, is fully aware of the flux in other map points and it deliberately minimizes the effect of these fluxes in the map point under consideration. Your second comment is completely to the point! Lack of time prevented me to mention this problem in my presentation; it is considered in the paper.

Comment J.E.B. PONSONBY
You said that one dimensional time series analysis and two dimensional imaging are analogous. You said that the Wiener-Khintchine theorem becomes the van Cittert-Zernike theorem, etc. You then discussed the time-series analysis of autoregressive processes with the implication that there is an analogy in 2 dimensions. Can you give examples of 2-dimensional autoregressive processes?
Reply C. van SCHOONEVELD
Yes and no. In my paper I give the example of a causal 2-dim. AR process, where the corner of a rectangular block of data can be predicted from the other points in the block. This is a direct analogy with the 1-dim. case. The trouble is, that one can imagine many other prediction schemes in 2 dimensions, most of them non-causal. Whether such processes do indeed exist, and whether they lead to modified MEM algorithms, is not yet completely clear.