Multichannel Maximum Entropy Spectral Analysis

In order to derive the multichannel maximum entropy spectral analysis equations, we must first define the entropy per time step of a set of stationary time series. If our N time series are statistically independent and gaussianly distributed, then the entropy per sample of each channel is just the integral of the logarithm of its spectrum and the entropy per step for all N time series is the sum of the entropies. Thus, the multichannel entropy for independent channels is

$$\sum_{n=1}^{N} \int \log P_n(f)\, df = \int \log \prod_{n=1}^{N} P_n(f)\, df = \int \log[\det(P(f))]df , \quad (1)$$

where $P(f)$ is the multichannel power spectrum matrix. In this case $P(f)$ is N by N and diagonal with its diagonal elements being given by $P_n(f)$, $n=1$ to N . We will now show that the integral of the logarithm of the determinant of $P(f)$ is a reasonable definition for the entropy of a general multichannel spectrum.

Suppose we have a set of N gaussianly distributed random variables v_n , with a general positive definite covariance matrix R . If the variables were independent, i.e., if R were diagonal, then the entropy of the set of variables would be the sum of the logarithms of their variances, or equivalently, the logarithm of the determinant of R . However, if the random variables are not independent, we can change them to a new set which are independent by using prediction error filters. For example, we can make our new set of variables be: v_N ; the error in predicting v_{N-1} from v_N ; the error in predicting v_{N-2}

from v_{N-1} and v_N ; etc., until we finally get for our last new variable the error in predicting v_1 from all the other variables. This set of prediction error variables will be independent and their entropy has been defined. It is seen that this procedure is analogous to that used to convert the correlated variables in a colored time series to the independent variables of a whitened time series. In matrix form, our "whitening" of the N by N covariance matrix looks like

$$U^T R U = D$$

where D is diagonal with positive elements d_n and U is the solution of the equation below (shown explicitly for N=4)

$$RU \equiv R \begin{bmatrix} 1 & 0 & 0 & 0 \\ * & 1 & 0 & 0 \\ * & * & 1 & 0 \\ * & * & * & 1 \end{bmatrix} = \begin{bmatrix} d_1 & * & * & * \\ 0 & d_2 & * & * \\ 0 & 0 & d_3 & * \\ 0 & 0 & 0 & d_4 \end{bmatrix} ,$$

where * indicates an element to be solved for. The first column of U is the filter for obtaining the least mean square error in predicting v_1 from v_n , n=2 to N , etc. We note that $|U| = 1$, so that $|D| = |R|$. Thus, the logarithm of the determinant of the covariance is invariant under the U transformation. This should seem reasonable since the unity weight on the predicted variable of the P.E.F. does not gain up the variable. Thus, since the U matrix is reversible and does not scale the variables, one can believe that it should not change the entropy. Using this result, if the multichannel power spectrum matrix is not diagonal but is constant with frequency, its entropy will be given by (1). However, we want to show that (1) is correct in general and to do this we call on multichannel prediction error filters. An M long multichannel P.E.F. has a matrix z transform of

$$F(z) = I + A_1 z + A_2 z^2 + \cdots + A_M z^M$$

where the leading matrix is the identity matrix and A_m , n=1 to M , are N by N matrices. The M.P.E.F. actually consists of N filters, where the nth filter predicts the next point on the nth channel from the past multichannel data. Because of the unity weight on the predicted point, the variables are not scaled and the entropy should not be changed. The inverse of the M.P.E.F. is also physically realizable and has the form

$$I + B_1 z + B_2 z^2 + \cdots \quad .$$

Just like the single channel P.E.F., the multichannel P.E.F. can "whiten" a multichannel time series, that is, convert the power spectral matrix to a constant matrix (but not necessarily a diagonal matrix). Likewise, since the multichannel P.E.F. is minimum phase, we find that the z polynomial given by the determinant of $F(z)$ has all of its roots outside the unit circle. Thus, using the single channel theorem, we have that

$$\int \log \; |F(z)| \; df = 0 \quad .$$

Now, if we have a multichannel time series with a constant spectral matrix P and filter it with $F(z)$, we obtain a new spectral matrix given by $F^{\dagger}(z) \; P \; F(z)$. However, we note that

$$\int \log |F^{\dagger}(z)\ P\ F\ (z)|\ df = \int \log |F^{\dagger}(z)|\ df + \int \log |P|\ df$$

$$+ \int \log |F(z)|\ df = \int \log |P|\ df \ .$$

Thus, the multichannel spectrum $F^{\dagger}(z)\ P\ F(z)$ has the same entropy as P . Since we can in general generate any multichannel spectrum by applying a M.P.E.F. to a constant spectral matrix, we can now see that (1) makes sense as the definition of entropy for any multichannel spectrum.

The constraint or measurement equations that we shall deal with are

$$\int P(z)\ z^m\ df = \Phi(m)\ , \quad m = -M \ \text{to} \ +M \tag{2}$$

where $\Phi(m)$ will be recognized as the N by N cross-correlation matrix at lag m of the multichannel time series. We note that (2) consists of $(2M+1)N^2$ equations. We thus need $(2M+1)N^2$ Lagrange multipliers, $\Gamma_{ij}(m)$, where i, j specify the matrix element in the mth equation. Our variation is thus taken over all N^2 functions in the spectral matrix $P(z)$ to give

$$\delta \int \left\{ \log |P(z)| - \sum_{i,j,m} \lambda_{ij}(m)\ [\ P_{ij}(z)\ z^m - \Phi_{ij}(m)\ ] \right\} df = 0\ , \text{ or}$$

$$\int \sum_{i,j} \left[\frac{Q_{ij}(z)}{|P(z)|} - \sum_{m=-M}^{+M} \lambda_{ij}(m)\ z^m \right] \delta P_{ij}(z)\ \ df = 0$$

where Q_{ij} is the cofactor of $P_{ij}(z)$ in the $P(z)$ matrix. Thus

$$\frac{Q_{ij}(z)}{|P(z)|} = P_{ij}^{-1}\ (z)\ \ ,$$

and we have

$$P_{ij}^{-1}(z) = \sum_{m=-M}^{+M} \lambda_{ij}(m)\, z^m \quad , \quad \text{or}$$

$$P^{-1}(z) = \sum_{m=-M}^{+M} \lambda(m)\, z^m \qquad \text{where}$$

$\lambda(m)$ is the matrix of the mth Lagrange multipliers. We now assert that if our constraint equations (2) are consistent, then it must be possible to write

$$\sum_{m=-M}^{+M} \lambda(m)\, z^m = F^{\dagger}(z) S\, F(z) \; ,$$

where $F^{\dagger}(z) = F_0 + F_1 z + F_2 z^2 + \cdots + F_M z^M$, $F_0 \equiv I$, is an Mth order M.P.E.F. and S is a constant power density matrix. We then have

$$P^{-1}(z) = F^{\dagger}(z)\; S\; F(z) \; , \quad \text{or}$$

$$F^{\dagger -1}(z)\; P^{-1}(z) = S\; F(z) \; , \text{or}$$

$$P(z)\; F^{\dagger}(z) = F^{-1}(z)\; S^{-1} \; .$$

Since $F(z)$ is minimum phase, $F^{-1}(z)$ contains no negative powers of z . Thus, the left hand side also cannot contain negative powers of z . Since

$$P(z) = \sum_{-\infty}^{+\infty} \Phi(m)\, z^m \quad \text{and} \quad F^{\dagger}(z) = \sum_{m=o}^{M} F_m\, z^{-m} \; ,$$

equality of the zero th power of z gives

$$\sum_{m=o}^{M} \Phi(m)\, F_m = S^{-1}$$

and equality of the rth power of z when r is negative gives

$$\sum_{m=o}^{M} \Phi(m+r)\, F_m = 0 \ .$$

Thus, specializing to the case where $M=3$, we have the multichannel prediction error filter equations

$$\begin{bmatrix} \Phi(0) & \Phi(1) & \Phi(2) & \Phi(3) \\ \Phi(-1) & \Phi(0) & \Phi(1) & \Phi(2) \\ \Phi(-2) & \Phi(-1) & \Phi(0) & \Phi(1) \\ \Phi(-3) & \Phi(-2) & \Phi(-1) & \Phi(0) \end{bmatrix} \begin{bmatrix} I \\ F_1^T \\ F_2^T \\ F_3^T \end{bmatrix} = \begin{bmatrix} S^{-1} \\ 0 \\ 0 \\ 0 \end{bmatrix} \ . \qquad (3)$$

We note that the square matrix in (3) is made up of our measured cross-correlation values and that we can solve for F_n and S from (3). The maximum entropy solution is then given by

$$P(f) = F^{-1}(z)\ S^{-1}\ F^{\dagger-1}(z) \ .$$