## II-A. SOLUTION OF THE VARIATIONAL FORMULATION

The problem is to find the real positive function $P(f)$ which maximizes

$$\int_{-W}^{W} \ln P(f) \; df \quad , \tag{II-1}$$

under the constraint equations

$$R(n) = \int_{-W}^{+W} P(f) \; e^{i2\pi fn\Delta t} \, df \; , \quad ( - N \leq n \leq N ) \quad . \tag{II-2}$$

### 1. The Functional Form of the Maximum Entropy Spectrum

One approach to solving this problem is to use Lagrange multipliers. However, since that approach will be used for more general cases later on, we shall use a different approach here. We begin by explicitly requiring that $P(f)$ be expressible in terms of a fourier series, i.e.,

$$P(f) = \frac{1}{2W} \sum_{n=-\infty}^{+\infty} R(n) \; e^{-i2\pi fn\Delta t} \quad . \tag{II-3}$$

Our constraint equations will be automatically satisfied by requiring that the $R(n)$ in (II-2) and (II-3) be the same for $-N \leq n \leq N$ . Substituting (II-3) into (II-1), we get

$$\int_{-W}^{W} \ln \left[ \frac{1}{2W} \sum_{n=-\infty}^{+\infty} R(n) \; e^{-i2\pi fn\Delta t} \right] df \; . \tag{II-4}$$

Taking the partial derivative of (II-4) with respect to $R(s)$ , where $|s| > N$ , we require that

$$\int_{-W}^{W} \left[ \frac{1}{2W} \sum_{n=-\infty}^{+\infty} R(n) \; e^{-i2\pi fn\Delta t} \right]^{-1} \frac{1}{2W} e^{-i2\pi fs\Delta t} \; df$$

$$= \frac{1}{2W} \int_{-W}^{W} P^{-1}(f) \; e^{-i2\pi fs\Delta t} \; df = 0 \quad , \quad |s| > N \; . \tag{II-5}$$

If we expand $P^{-1}(f)$ in a fourier series,

$$P^{-1}(f) = \sum_{n=-\infty}^{+\infty} \lambda_n e^{-i2\pi fn\Delta t} ,$$

then (II-5) shows that $\lambda_n = 0$ for $|n| > N$ . Thus, we find that the functional form of $P(f)$ is given by

$$P(f) = \frac{1}{\sum_{s=-N}^{N} \lambda_s e^{-i2\pi fs\Delta t}} . \qquad \text{(II-6)}$$

This is the same equation that would be derived from the Lagrange multiplier approach, where the $\lambda$'s would be the Lagrange multipliers.

The next step in our solution procedure is to determine values for the $\lambda$'s in (II-6) such that the constraint equations (II-2) will be satisfied. We shall do this in two different ways.
The first method will use analytic integration around the unit circle in the complex $z$ plane. The other method will use a simple z-transform argument.

2. The Analytic Integration Derivation

The analytic integration derivation begins by substituting (II-6) into (II-2) to get the 2N+1 equations,

$$\int_{-W}^{+W} \frac{e^{i2\pi fn\Delta t}}{\sum_{s=-N}^{N} \lambda_s e^{-i2\pi fs\Delta t}} df = R(n) , \quad ( -N \leq n \leq N ) . \qquad \text{(II-7)}$$

We now convert to $z$ transform notation by setting $z = e^{-i2\pi f\Delta t}$ .
Then, since

$$dz = -i2\pi\Delta t \; e^{-i2\pi f\Delta t} \; df = -i2\pi\Delta t \; z \; df \quad , \text{ or}$$

$$df = \frac{-dz}{i2\pi\Delta t \; z} \quad , \text{ equation II-7 becomes}$$

$$\int_{-W}^{+W} \frac{z^{-n} \; df}{\sum_{s=-N}^{+N} \lambda_s \; z^s} = \frac{1}{2\pi i\Delta t} \oint \frac{z^{-n-1} \; dz}{\sum_{s=-N}^{+N} \lambda_s \; z^s} = R(n) \; , \qquad \text{(II-8)}$$

where the contour integral is around the unit circle in the counter-clockwise direction. Noting that since we are requiring $P(f)$ to be real and positive for $|z| = 1$ , we see from (II-6) that it must be possible to write

$$\sum_{s=-N}^{N} \lambda_s \; z^s = [P_N\Delta t]^{-1}[\; 1 + a_1 \; z + \ldots + a_N \; z^N \;][\; 1 + a_1^* \; z^{-1} + a_N^* \; z^{-N} \;]$$

$$= [P_N\Delta t]^{-1} \sum_{s=0}^{N} a_s \; z^s \sum_{s=0}^{N} a_s^* \; z^{-s} \quad , \qquad \text{(II-9)}$$

where $P_N > 0$ and $a_0 = 1$ . Furthermore, all of the roots of the first polynomial in $z$ can be chosen to lie outside the unit circle and thus all of the roots of the second polynomial will lie inside the unit circle. Substituting (II-9) into (II-8) gives,

$$\frac{P_N}{2\pi i} \oint \frac{z^{-n-1} \qquad dz}{\sum_{s=0}^{N} a_s \; z^s \; \sum_{s=0}^{N} a_s^* \; z^{-s}} = R(n) \; , \quad ( -N \leq n \leq N ) \; . \qquad \text{(II-10)}$$

We now form the summations

$$\sum_{n=0}^{N} a_n^* R(n-r) = \frac{P_N}{2\pi i} \oint \frac{z^{r-1} \sum_{n=0}^{N} a_n^* z^{-n}}{\sum_{s=0}^{N} a_s z^s \sum_{s=0}^{N} a_s^* z^{-s}} dz =$$

$$\frac{P_N}{2\pi i} \oint \frac{z^{r-1} dz}{\sum_{s=0}^{N} a_s z^s} , \quad r \geqslant 0 . \qquad \text{(II-11)}$$

Remembering that the denominator polynomial is analytic on and inside the unit circle, we see that for $r \geqslant 1$, the integrand in (II-11) is analytic on and inside our contour of integration and thus the integral is zero. To handle the $r=0$ case, we note that one form of Cauchy's integral formula is

$$\frac{1}{2\pi i} \oint \frac{f(z)}{z} dz = f(0) ,$$

where $f(z)$ is analytic on and inside the contour of integration, and where the contour encloses the origin. Using this formula and remembering that $a_0 = 1$, we see that for $r=0$, (II-11) is equal to $P_N$. Taking the complex conjugate of these equations, we have

$$\sum_{n=0}^{N} R(r-n)\, a_n = P_N \quad \text{for} \quad r=0 \text{, and}$$

$$\sum_{n=0}^{N} R(r-n)\, a_n = 0 \quad \text{for} \quad r \geqslant 1 \quad . \qquad \text{(II-12)}$$

3. <u>The Z-Transform Derivation</u>

The z-transform derivation of (II-12) starts by substituting (II-3) and (II-9) into (II-6) to get

$$P(f) = \frac{P_N \Delta t}{\sum_{s=0}^{N} a_s z^s \sum_{n=0}^{N} a_n^* z^{-n}} = \frac{1}{2W} \sum_{r=-\infty}^{+\infty} R(r) z^r . \qquad \text{(II-13)}$$

Using the fact that $2W \Delta t = 1$ and multiplying through by the maximum phase factor, we get

$$\frac{P_N}{\sum_{s=0}^{N} a_s z^s} = \sum_{n=0}^{N} a_n^* z^{-n} \sum_{r=-\infty}^{+\infty} R(r) z^r$$

$$= \sum_{n=0}^{N} a_n^* z^{-n} \sum_{r=-\infty}^{+\infty} R(n-r) z^{n-r} = \sum_{r=-\infty}^{+\infty} \left[ \sum_{n=0}^{N} a_n^* R(n-r) \right] z^{-r} . \qquad \text{(II-14)}$$

Noting that since the first expression in (II-14) is minimum phase and is thus analytic on and inside the unit circle, its z transform involves only non-negative powers of $z$. Also since $a_0=1$, the coefficient of $z^0$ is $P_N$. Thus, equating coefficients of the first and last expressions in (II-14), and taking complex conjugates, we again derive the set of equations (II-12).

4. The Prediction Error Filter Equation

Equations (II-12) for $0 \le r \le N$ can be written in matrix form as

$$\begin{bmatrix} R(0) & R(-1) & & R(-N) \\ R(1) & R(0) & & R(1-N) \\ & & & \\ R(N) & R(N-1) & & R(0) \end{bmatrix} \begin{Bmatrix} 1 \\ a_1 \\ \\ a_N \end{Bmatrix} = \begin{Bmatrix} P_N \\ 0 \\ \\ 0 \end{Bmatrix} . \qquad \text{(II-15)}$$

Equation (II-15) is the well-known equation for finding the Nth order prediction error filter and will be discussed in that context in section II-C. Its solution for $P_N$ and $a_n$, $(n = 1 \text{ to } N)$, is easily accomplished by using a modern version of the Levinson algorithm. Once the $a_n$ and $P_N$ are found, they are substituted into (II-13) to obtain the maximum entropy spectrum. We shall now discuss and derive the modern version of the Levinson algorithm since it is of central importance in the study of autocorrelation functions.

5. The Modern Levinson Algorithm

Norman Levinson's algorithm for solving (II-15) was presented in an appendix to Norbert Wiener's book, Extrapolation, Interpolation and Smoothing of Stationary Time Series.[5] This algorithm takes advantage of the Toeplitz form of the matrix in (II-15) which contains the autocorrelation lag values. The number of arithmetical operations required to solve a general set of $N$ linear simultaneous equations in $N$ unknowns is on the order of $N^3$. The Levinson algorithm is on the order of $N^2$ operations and is thus appreciably faster than the general gaussian elimination procedure. In addition, the Levinson algorithm only requires memory storage on the order of $N$, compared to an $N^2$ order for the general case. Another advantage of the Levinson algorithm is that it is recursive. That is, the solution to the $N+1^{th}$ set of equations is obtained from the solution of the $N^{th}$ set of equations. This means that in solving the $N^{th}$ order set of equations, all lower order solutions are obtained and that higher order solutions can be obtained without wasted effort.

The original Levinson algorithm requires the equivalent of three vector dot products per recursion. One of these is used to calculate the value of $P_N$ . Since $P_N$ may become quite small, the accumulation of round-off error can make the algorithm numerically unstable (see Brouwer,[6] 1971; Pagano,[7] 1972). Fortunately, in 1961 Burg discovered that $P_N$ can be obtained directly from $P_{N-1}$ by a single accurate operation. Thus, the equivalent of only two vector dot products is required per recursion, speeding up the algorithm by a third as well as improving the numerical stability. While this discovery was incorporated in the papers by Robinson and Wiggins,[8] they did not explicitly note that the modern algorithm is different from the one that Levinson derived. Unfortunately, Brouwer,[6] 1971 and Pagano,[7] 1972 were unaware of this difference.

In deriving the modern algorithm, we shall assume that the $N$ by $N$ Toeplitz submatrix is positive definite and that the full $N+1$ by $N+1$ Toeplitz matrix is at least non-negative definite. This will be true in all practical cases with which we shall be concerned. Then (II-15) will always have a solution and this solution will be unique. Thus, if we can find a solution to (II-15) in which the first element of the left hand column vector is unity and the last $N$ elements of the right hand column vector are zero, we have found the one and only solution to (II-15).

The fact that $P_N$ is real is proven by noting that the autocorrelation matrix is hermitian and that premultiplying (II-15) by the row vector $(1, a_1^*, a_2^*, \ldots, a_N^*)$ gives $P_N$. Also, since multiplying the second column of the autocorrelation matrix by $a_1$ and then adding it to the first column, etc., does not change the determinant of the matrix, we see that this determinant is equal to $P_N$ times the determinant of the N by N submatrix. Thus $P_N > 0$ if the N+1 matrix is positive definite and $P_N = 0$ if the N+1 matrix is singular.

Starting with the solution of the set of N equations,

$$\begin{bmatrix} R(0) & R(-1) & & R(1-N) \\ R(1) & R(0) & & R(2-N) \\ & & & \\ R(N-1) & R(N-2) & & R(0) \end{bmatrix} \begin{Bmatrix} 1 \\ b_1 \\ \\ b_{N-1} \end{Bmatrix} = \begin{Bmatrix} P_{N-1} \\ 0 \\ \\ 0 \end{Bmatrix} ,$$

the algorithm for solving the set of N+1 equations is most easily developed by studying the matrix equation

$$\begin{bmatrix} R(0) & R(-1) & & R(1-N) & R(-N) \\ R(1) & R(0) & & R(2-N) & R(1-N) \\ & & & & \\ R(N-1) & R(N-2) & & R(0) & R(-1) \\ R(N) & R(N-1) & & R(1) & R(0) \end{bmatrix} = \left\{ \begin{Bmatrix} 1 \\ b_1 \\ \\ b_{N-1} \\ 0 \end{Bmatrix} + c_N \begin{Bmatrix} 0 \\ b_{N-1}^* \\ \\ b_1^* \\ 1 \end{Bmatrix} \right\} = \left\{ \begin{Bmatrix} P_{N-1} \\ 0 \\ \\ 0 \\ \Delta_N \end{Bmatrix} + c_N \begin{Bmatrix} \Delta_N^* \\ 0 \\ \\ 0 \\ P_{N-1} \end{Bmatrix} \right\}. \tag{II-16}$$

In looking at this equation, we note that the second column vector on both sides of (II-16) is the simple complex conjugate reverse of the first column. The reason that the equation is valid for these complex conjugate reverse vectors is that the first and last rows of the autocorrelation matrix are complex conjugate reverses, the 2nd and next to last rows are complex conjugate reverses, etc. From inspection, we see that (II-16) defines $\Delta_N$ to be

$$\Delta_N = \sum_{n=0}^{N-1} R(N-n)\, b_n \quad , \tag{II-17}$$

where $b_0 = 1$ . We now form the Nth order prediction error filter equation from (II-16) by specifying the value of the Nth reflection coefficient, $c_N$ , to be

$$c_N = - \Delta_N / P_{N-1} \quad ,$$

so that

$$\Delta_N + c_N P_{N-1} = 0 \quad . \tag{II-18}$$

This equation for $c_N$ will always have a solution since we have assumed $P_{N-1} > 0$ . By using this value of $c_N$ to combine the two column vectors on both sides of (II-16), we note that the left side column vector starts with unity and the right side column vector is all zeros below the first element. Thus, we have indeed formed (II-15) from (II-16).

The Nth order prediction error filter is thus generated from $c_N$ and the N-1 th order filter by

$$\begin{Bmatrix} 1 \\ a_1 \\ \\ a_{N-1} \\ a_N \end{Bmatrix} = \begin{Bmatrix} 1 \\ b_1 \\ \\ b_{N-1} \\ 0 \end{Bmatrix} + c_N \begin{Bmatrix} 0 \\ b^*_{N-1} \\ \\ b^*_1 \\ 1 \end{Bmatrix} \quad . \tag{II-19}$$

To obtain the value of $P_N$, the Levinson algorithm multiplied the Nth order prediction error filter through the top row of the autocorrelation matrix. However, we see from (II-16) that $P_N = P_{N-1} + c_N \Delta_N^*$, or using (II-18),

$$P_N = P_{N-1} ( 1 - c_N c_N^* ) = P_{N-1} ( 1 - |c_N|^2 ) . \qquad \text{(II-20)}$$

To initiate this recursive algorithm, we start with $P_0 = R(0)$ and the zero$^{th}$ order prediction error filter, which is simply the one point filter with unity weight.

One should note that (II-19) is reversible if $|c_N| \neq 1$. That is, given the Nth order prediction error filter, one can obtain $c_N$ and the N-1th order prediction error filter if $a_N = c_N$ does not have unit magnitude. To prove this, we note from (II-19) that

$$a_s = b_s + c_N b_{N-s}^* \quad \text{and}$$

$$a_{N-s} = b_{N-s} + c_N b_s^* .$$

Taking the complex conjugate of the second equation, we find in matrix form that

$$\begin{Bmatrix} a_s \\ a_{N-s}^* \end{Bmatrix} = \begin{bmatrix} 1 & c_N \\ c_N^* & 1 \end{bmatrix} \begin{Bmatrix} b_s \\ b_{N-s}^* \end{Bmatrix} , \quad \text{or}$$

$$\begin{Bmatrix} b_s \\ b_{N-s}^* \end{Bmatrix} = \frac{1}{(1-|c_N|^2)} \begin{bmatrix} 1 & -c_N \\ -c_N^* & 1 \end{bmatrix} \begin{Bmatrix} a_s \\ a_{N-s}^* \end{Bmatrix} . \qquad \text{(II-21)}$$

This inverse of course exists only if $|c_N| \neq 1$. We can note that (II-21) is valid even when $s = N-s$.

## 6. General Comments

Before concluding this section, a few general comments about the solution to the maximum entropy variational principle are in order. First, it is implied that the set of measurements (II-2) are consistent with some positive function P(f) . In the derivation used in this section, (II-3) explicitly requires that the R(n) , $|n| \leq N$ , be such that some positive function P(f) exists that satisfies (II-3). We shall show in the next section that this will be true if and only if the N+1 by N+1 autocorrelation matrix is positive definite. Secondly, it is assumed that the constraint equations set a finite upper bound on (II-1). With no upper bound, there would be no finite solution to setting the partial derivatives equal to zero. It is clear, however, that constraining the total power to be R(0) places an upper bound on (II-1) because of the convexity of the logarithmic function. Third, equations (II-10) and (II-13) both assume that the functional form involving the a's is general enough to satisfy any consistent set of R(n) , $n \leq N$ . This would not be true if the two z polynomials appeared in the numerator. The fact that (II-10) and (II-13) are general enough is the result of being derived from a variational principle in which the P(f) solution is constrained to be positive. Fourth, from the derivation, it is clear that the maximum entropy solution for any consistent set of constraint equations (II-2) will always exist and be unique.